DTEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE
ON DEFENSIVE TECHNOLOGIES

PROCEEDINGS

TOPICS

AERODYNAMICS AND FLIGHT DYNAMICS
AIRCRAFT
WEAPON SYSTEMS AND COMBAT VEHICLES
AMMUNITION AND ENERGETIC MATERIALS
INTEGRATED SENSOR SYSTEMS AND ROBOTIC SYSTEMS
TELECOMMUNICATION AND INFORMATION SYSTEMS
MATERIALS AND TECHNOLOGIES
QUALITY, STANDARDIZATION, METROLOGY, MAINTENANCE AND EXPLOITATION

OTEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

Republic of Serbia, Tara
October 21st - 24th, 2026

Miomir Vukobratović
1931 - 2012

Publisher
The Military Technical Institute
Ratka Resanovića 1, 11030 Belgrade

Publisher's Representative
Col Ivan Pokrajac, PhD (Eng)

Editors
Col Aleksandar Kari, PhD (Eng)
Prof Ljubica Radović, PhD (Eng)

Technical Editing
Aida Jeremić
Jelena Krsmanović

Printed by
The Military Technical Institute
Ratka Resanovića 1, 11030 Belgrade

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

NAUČNO-stručni skup iz oblasti odbramenbih tehnologija (12 ; 2026 ; Tara)
Proceedings [Elektronski izvor] / 12th International Scientific Conference on Defensive Technologies, OTEH 2026, Tara, 21-24 October 2026 ; [organized by] Military Technical Institute, Belgrade ; [editors Aleksandar Kari, Ljubica Radović]. - Belgrade : Military Technical Institute, 2026 (Beograd : Military Technical Institute) 1 USB memorija.

Sistemski zahtevi: Nisu navedeni. - Nasl. sa naslovne strane dokumenta. -
Bibliografija uz svaki rad.

ISBN-978-86-81123-95-9

1. Vojnotehnički institut (Beograd)
a) Војна техника - Зборници
b) Технички материјали – Зборници
c) Телекомуникациони системи - Зборници

12th INTERNATIONAL SCIENTIFIC CONFERENCE
ON DEFENSIVE TECHNOLOGIES

OTEH 2026

SUPPORTED BY

MINISTRY OF DEFENCE
www.mod.gov.rs

Organized by

MILITARY TECHNICAL INSTITUTE
1 Ratka Resanovića St., Belgrade 11000, SERBIA
www.vti.mod.gov.rs

SUPPORTED BY

MINISTRY OF SCIENCE TECHNOLOGICAL DEVELOPMENT AND INNOVATION
www.nitra.gov.rs

ORGANIZATIONAL COMMITTEE

Nenad Miloradović, PhD Eng, Acting Assistant Minister of Defense for Materiel Resources, ***President of the Organizational Committee***
Major General **Slavko Rakić**, PhD Eng, Head of the Department for Defense Technologies
colonel **Ivan Pokrajac**, PhD Eng, Director of the Military Technical Institute
Brigadier General **Stanisav Mijailović**, Head of the Department for Planning and Development - GS SAF
colonel **Dejan Stojković**, PhD Eng, Head of the Department for Strategic Planning
colonel **Vinko Žnidaršič**, PhD Eng, Department for Strategic Planning
Prof. **Boban Đorović**, PhD Eng, rector of the University of Defense
Major General **Željko Furundžić**, Head of the Military Academy
colonel **Zlatko Bogosavljević**, Head of the Department for International Military Cooperation
colonel **Saša Veselinović**, PhD Eng, Director of the Technical Test Center
colonel **Boriša Jovanović**, PhD Eng, Applied Mathematics and Electronics Center – GS SAF
colonel **Aleksandar Mitić**, BSc Eng, Director (authorized by the Minister of Defense) od the Directorate for Standardization, Codification and Metrology
lt col **Ljubo Kurćubić**, Director of the Military Quality Control
colonel **Saša Bakrač**, Phd Eng, Acting Director of the Military Geographical Institute
Prof. **Dragan Stanar**, PhD Eng., Strategic research Institute - MoD RS
Ana Marinković, Assistant Minister, Sector for Innovations, transfer of Technologies and Technological development, Ministry of Science, Technological Development and Innovations
Vladimir Popović, PhD Eng, Dean of the Faculty of Mechanical Engineering, University of Belgrade
Prof. **Zoran Čiča**, PhD Eng, Acting Dean of the Faculty of Electrical Engineering, University of Belgrade
Prof. **Mirjana Kijevčanin**, PhD Eng, Dean of the Faculty of Technology and Metallurgy, University of Belgrade
Boris Dumnić, PhD Eng, dean of the Faculty of Technical Sciences, University of Novi Sad
Prof. **Slobodan Savić**, PhD Eng, Dean of the Faculty of Engineering Sciences, University of Kragujevac
Prof. **Ivan Milićević**, PhD Eng, Acting Dean of the Faculty of Technical Sciences, Čačak, University of Kragujevac
Prof. **Goran Janevski**, PhD Eng, Dean of the Faculty of Mechanical Engineering, University of Niš
Prof. **Vladimir Ćirić**, PhD Eng, Dean of the Faculty of Electronic Engineering, University of Niš
Vladimir Kostić, BSc Eng, Director, IRITEL, Belgrade
Vladimir Cizelj, PhD Eng., Vlatacom Institute, Belgrade
Nikola Tomašević, PhD Eng., Director, Mihailo Pupin Institute
Jugoslav Petković, Director, "Yugoimport – SDPR", Belgrade
Jovanka Andrić, Director, "Krušik", Valjevo
Zoran Stefanović, Director, "Sloboda", Čačak
Predrag Ilić, Director "Milan Blagojević – namenska A.D.", Lučani
Dobrosav Andrić, Director, "Prvi partizan", Užice
Stanoje Biočanin, Director, "Prva Iskra-namenska", Barič
Saša Milošević, Director, "Zastava oružje", Kragujevac
Andrijana Jovanović, Director, "Prva Petoletka – Namenska", Trstenik
Stefan Đurić, Director, "Zastava TERVO", Kragujevac
Časlav Kljajić, Director, "PPT-TMO A.D.", Trstenik

SCIENTIFIC COMMITTEE

colonel **Aleksandar Kari**, PhD Eng, Military Technical Institute, Belgrade, Serbia, ***President of the Scientific Committee***
Prof. **Ljubica Radović**, PhD Eng, Military

Technical Institute, Belgrade, Serbia
colonel **Bojan Pavković,** PhD Eng, Research Associate, Military Technical Institute, Belgrade, Serbia,
colonel **Dejan Ivković**, PhD Eng, Research Associate, Military Technical Institute, Belgrade, Serbia,
Prof. **Taek Lyul Song**, PhD Eng, Hanyang University, South Korea
Evgeny V. Sudov, PhD Eng, R&D Applied Logistic Centre, Moscow, Russia
Stevan Berber, PhD Eng, Auckland University, New Zealand
Srećko Stopić, PhD Eng, RWTH University, Aachen, Germany
Zbysek Korecki, PhD Eng, University of Defence, Brno, Czech Republic
Radi Ganev, PhD Eng, University "Lyuben Karavelov", Sofia, Bulgaria
Prof. **Constanca Rigueiro**, PhD Eng, Polytechnic Institute Castelo Branco, Portugal
Prof. **Tomaž Vuherer**, PhD Eng, Faculty of Mechanical Engineering, Maribor, Slovenia
Prof. **Biljana Marković**, PhD Eng, Faculty of Mechanical Engineering, East Sarajevo, Republic of Srpska (Bosnia and Herzegovina)
Prof. **Milan Tica**, PhD Eng, Faculty of Mechanical Engineering, University of Banja Luka, Republic of Srpska (Bosnia and Herzegovina)
Prof. **Dženana Gačo**, PhD Eng, University of Bihać, Bosnia and Herzegovina
Prof. **Mersida Manjgo**, PhD Eng, Faculty of Mechanical Engineering Mostar, Bosnia and Herzegovina
Prof. **Kemal Delijić**, PhD Eng, Faculty of Metallurgy and Technology, Podgorica, Montenegro
colonel **Predrag Rakonjac**, PhD Eng, University of Defense, Belgrade, Serbia
colonel Prof. **Dragan Trifković**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof. **Slobodan Simić**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof. **Damir Jerković**, PhD Eng, Military Academy, Belgrade, Serbia
colonel **Branimir Krstić**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof. **Boban Pavlović**, PhD Eng, Military Academy, Belgrade, Serbia
colonel **Boban Bondžulić**, PhD Eng, Military Academy, Belgrade, Serbia
colonel **Milan Mihajlović**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof. **Milenko Andrić**, PhD Eng, Military Academy, Belgrade, Serbia
colonel **Dimitrije Bujaković**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof. **Zoran Bajić**, PhD Eng, Military Academy, Belgrade, Serbia
colonel Prof **Radovan Karkalić**, PhD Eng, Military Academy, Belgrade, Serbia
lieutenant colonel Prof. **Mihael Bučko**, PhD Eng, Military Academy, Belgrade, Serbia
colonel **Nenad Stefanović**, PhD Eng, Applied Mathematics and Electronics Center - GS SAF, Belgrade, Serbia
colonel **Saša Bakrač**, PhD Eng, Military Geographical Institute, Belgrade, Serbia
Jasmina Stevanović, PhD Eng, Institute of Chemistry, Technology and Metallurgy, Belgrade, Serbia
Srećko Manasijević, PhD Eng, Principal Research Fellow, Lola Institute, Belgrade, Serbia
Dragana Perić, PhD Eng, Vlatacom Institute, Belgrade, Serbia
Srećko Nijemčević, PhD Eng, Vlatacom Institute, Belgrade, Serbia
Ivan Vulić, PhD Eng, Vlatacom Institute, Belgrade, Serbia
Miroslav Sokić, PhD Eng, Principal Research Fellow, Institute for Technology of Nuclear and Other Raw Materials, Belgrade, Serbia
Prof. **Stevica Graovac**, PhD Eng, Mihailo Pupin Institute, Belgrade, Serbia
Nikola Zogović, PhD Eng, Principal Research Fellow**,** Mihailo Pupin Institute, Belgrade, Serbia
Violeta Nikolić, PhD Eng, Senior Research Assistant, Vinča Institute of Nuclear Sciences, Belgrade, Serbia
Ivana Perović, PhD Eng, Senior Research Assistant, Vinča Institute of Nuclear Sciences, Belgrade, Serbia
Nataša Lazarević, PhD Eng, Nuclear Facilities of Serbia, Belgrade, Serbia
Prof. **Vladimir Mladenović**, PhD Eng, Faculty of Engineering Sciences - Čačak, University of Kragujevac, Serbia
Dušan Arsić, PhD Eng, Faculty of Engineering Sciences, University of Kragujevac, Serbia
Prof. **Vladimir Arsoski**, PhD Eng, Faculty of Electrical Engineering, University of Belgrade, Serbia

Prof. **Dragan Milčić**, PhD Eng, Faculty of Mechanical Engineering, University of Niš, Serbia
Prof. **Dragan Šešlija**, PhD Eng, Faculty of Technical Sciences, University of Novi Sad, Serbia
Prof. **Predrag Elek**, PhD Eng, Faculty of Mechanical Engineering, University of Belgrade, Serbia
Prof. **Aleksandar Simonović**, PhD Eng, Faculty of Mechanical Engineering, University of Belgrade, Serbia
Prof. **Jelena Svorcan**, PhD Eng, Faculty of Mechanical Engineering, University of Belgrade, Serbia
Prof. **Ivan Kostić**, PhD Eng, Faculty of Mechanical Engineering, University of Belgrade, Serbia
Prof. **Aleksandar Marinković**, PhD Eng, Faculty of Technology and Metallurgy, University of Belgrade, Serbia
Prof. **Dragomir Glišić**, PhD Eng, Faculty of Technology and Metallurgy, University of Belgrade, Serbia
Marija Samardžić, PhD Eng, Military Technical Institute, Belgrade, Serbia
Jelena Gržetić, PhD Eng, Military Technical Institute, Belgrade, Serbia
Prof. **Maja Vitorović-Todorović**, PhD Eng, Military Technical Institute, Belgrade, Serbia
Miloš Pavić, PhD Eng, Principal Research Fellow, Military Technical Institute, Belgrade, Serbia
lieutenant colonel **Saša Brzić**, PhD Eng, Military Technical Institute, Belgrade, Serbia
Danica Bajić, PhD Eng, Military Technical Institute, Belgrade, Serbia
colonel **Rade Pavlović**, PhD Eng, Military Technical Institute, Belgrade, Serbia
lieutenant colonel **Boban Sazdić-Jotić**, PhD Eng, Principal Research Fellow, Military Technical Institute, Belgrade, Serbia
Miodrag Milenković-Babić, PhD Eng, Principal Research Fellow, Military Technical Institute, Belgrade, Serbia
Srđa Perković, PhD Eng, Principal Research Fellow, Military Technical Institute, Belgrade, Serbia

OTEH 2026 SECRETARIAT

Elida Vasiljević, Military Technical Institute, secretary
Mirjana Krstović, Military Technical Institute, assistant secretary
col Branko Medan, Department of Defense Technologies
lt col Slobodan Malbašić, Department of Defense Technologies
lt col Nebojša Grebović, Military Technical Institute
lt col Igor Đorić, Military Technical Institute
Jelena Pavlović, Military Technical Institute
Jelena Krsmanović, Military Technical Institute
Katarina Stanisavljević, Military Technical Institute
Katarina Nestorović, Military Technical Institute
Aleksandar Ćitić, Military Technical Institute
Ljiljana Obradović, Military Technical Institute
Suzana Vuković, Military Technical Institute
Dobrivoje Granžan, Military Technical Institute
Aida Jeremić, Military Technical Institute

PREFACE

Military Technical Institute, the first and the largest military scientific-research institution in the Republic of Serbia with over 77 years long tradition, has been traditionally organizing the OTEH scientific conference, devoted to defense technologies. The Conference is supported by the Ministry of Defense and it takes place every second year.

Its aim is to gather scientists and engineers, researchers and designers, manufactures and university professors in order to exchange ideas and to develop new relationships.

The 12th International Scientific Conference OTEH 2026 is scheduled as follows: lecture on the occasion of "Miomir Vukobratović", given by Aleksandar Rodić, PhD Eng, and two plenary lectures: given by representatives of DRS RADA Technologies, and presented by Velibor Ilić, PhD Eng "Artificial Intelligence from Perception to Autonomous Action: Applications, Opportunities and Challenges" .

The papers which will be presented at the Conference have been classified into the following topics:
− Aerodynamics and Flight Dynamics
− Aircraft
− Weapon Systems and Combat Vehicles
− Ammunition and Energetic Materials
− Integrated Sensor Systems and Robotic Systems
− Telecommunication and Information Systems
− Materials and Technologies
− Quality, Standardization, Metrology, Maintenance and Exploitation.

The Proceedings contain 160 reviewed papers which have been submitted by the authors from 13 different countries. I would also like to emphasize that we have 15 papers with authors from abroad. The quality of papers accepted for publication achieved very high standard. I expect stimulated discussion on many topics that will be presented online, during two days of the Conference.

On behalf of the organizer I would like to thank all the authors and participants from abroad, as well as from Serbia, for their contribution and efforts which made this Conference possible and successful.

I would also like to thank the Ministry of Science, Technological Development and Innovations of the Republic of Serbia for its financial support.

Finally, dear guests and participants of the Conference, I would like to wish you a pleasant and successful work during the Conference. I am looking forward to see you again at the Conference. All the best and stay healthy.

Serbia, Tara, October, 2026

Col. Aleksandar Kari PhD Eng
President of the Scientific Committee
OTEH 2026

TABLE OF CONTENTS

1. SECTION : AERODYNAMICS AND FLIGHT DYNAMICS - AFD

2 **COMPARATIVE STUDY OF HYBRID DOLPHIN AND OTHER BIONIC AIRFOILS FOR DEFENSE APPLICATIONS,** Zorana Z. Dančuo, Ivan A. Kostić, Olivera P. Kostić, Aleksandar Č. Bengin, Goran S. Vorotović
10.5937/OTEH26001B

8 **INFLUENCE OF UNDERWING PAYLOAD ON THE FLIGHT PERFORMANCE OF A TACTICAL UAV,** Branislav Ostojić, Miodrag Milenković-Babić, Aljoša Ljuboja, Ognjen Peković
10.5937/OTEH26002O

14 **SYSTEMATIC MODELING AND EXPERIMENTAL EVALUATION OF BLDC MOTOR-PROPELLER SYSTEMS IN MULTICOPTERS,** Aljoša Ljuboja, Milenko Trifković, Branislav Ostojić, Toni Ivanov
10.5937/OTEH26003L

21 **SIMULATING SHOCK-WAVE/BOUNDARY-LAYER INTERACTIONS,** Jelena Svorcan, Dragoljub Tanović, Milica Ivanović
10.5937/OTEH26004S

27 **HIGH-PRESSURE JET SUPPLY SYSTEM FOR POWERED WIND TUNNEL MODELS: CALIBRATION AND PERFORMANCE,** Biljana Ilić, Marija Samardžić, Božidar Simović, Isidora Stanković, Ognjen Babić, Goran Ocokoljić
10.5937/OTEH26005I

33 **COMPARISON OF MISSILE AERODYNAMIC DATA OBTAINED IN WIND TUNNEL TESTS USING TWO INTERNAL STRAIN GAUGE BALANCES,** Marija Samardžić, Biljana Ilić, Isidora Stanković, Božidar Simović, Ognjen Babić, Goran Ocokoljić
10.5937/OTEH26006S

39 **SIDE JET INTERACTION WITH SUBSONIC, TRANSONIC AND SUPERSONIC CROSSFLOW: SURFACE PRESSURE DISTRIBUTION,** Božidar Simović, Biljana Ilić, Marija Samardžić, Isidora Stanković, Ognjen Babić, Goran Ocokoljić
10.5937/OTEH26007S

2. SECTION : AIRCRAFT - A

44 **SM2-M UAV AND Q GROUND CONTROL: SYNERGY OF AUTONOMOUS GUIDANCE AND REAL-TIME MISSION MONITORING IN MODERN MILITARY OPERATIONS,** Miloš Ugrenović
10.5937/OTEH26008U

48 **AIRCRAFT LIFE EXTENSION THROUGH INTEGRATED STRUCTURAL ASSESSMENT AND SYSTEM REVALIDATION,** Marina Ostojić, Tamara Nikolić, Marko Prošić
10.5937/OTEH26009P

54 **DRONE MILITARY TESTING,** Stevan Jovičić, Danilo Vukosavljević, Nikola Zdravković, Jelena Šobot
10.5937/OTEH26010J

59 **CRITICAL LOAD CALCULATION FOR SOME KIND OF NONUNIFORM STRUTS,** Bogdan S. Bogdanović, Tonko A. Mihovilović
10.5937/OTEH26011B

65 **PLATE SPRING-BASED NEGATIVE STIFFNESS ROTARY MECHANISM FOR PASSIVE ENERGY BALANCING OF MORPHING WINGS,** Mašan Vučinić, Toni Ivanov, Jelena Svorcan
10.5937/OTEH26012V

72 **DESIGN AND VALIDATION OF TOPOLOGY-OPTIMIZED 3D-PRINTED QUADCOPTER ARMS,** Milan Z. Rakić, Toni D. Ivanov
10.5937/OTEH26013R

77 **PROPELLER-MOTOR MATCHING FOR MINI-UAV: THRUST AND EFFICIENCY ESTIMATION BY USING THE LY-MICRO-MAX TEST STAND,** Miodrag Milenković-Babić, Branislav Ostojić, Aljoša Ljuboja, Mario Silvagni
10.5937/OTEH26014M

82 **ESSENTIAL TECHNICAL KNOWLEDGE AND SKILLS FOR FUTURE MILITARY AND CIVIL PILOTS IN THE REPUBLIC OF SERBIA,** Slaviša Vlačić, Aleksandar Knežević, Nebojša Radojković
10.5937/OTEH26015V

3. SECTION : WEAPON SYSTEMS AND COMBAT VEHICLES – WSCV

89 **A BALLISTIC MODULE ARCHITECTURE FOR MORTAR FIRE CONTROL SYSTEMS,** Miloš Marković, Andrija Ostojić, Dejan Jevtić
10.5937/OTEH26016M

96 **EXPERIMENTAL STUDY OF FRACTURED PROPELLANT INFLUENCE ON THE IMPULSE ROCKET MOTOR SAFETY,** Katarina Nestorović, Saša Živković, Marina Simović – Pavlović, Veselin Živanović, Jelena Mojsilović, Darko Janković
10.5937/OTEH26017N

101 **INTEGRATED TECHNICAL MODEL OF DETECTION AND RISK MANAGEMENT OF EXPLOSIVE REMNANTS OF WAR - CASE STUDU SMEDEREVO FORTRESS,** Anita Klikovac, Sanela Veljković, Marina Dabetić
10.5937/OTEH26018K

109 **COMPARATIVE ANALYSIS OF THEORETICAL AND EXPERIMENTAL MUZZLE VELOCITIES FOR THE EVALUATION OF POSSIBLE BARREL WEAR FOR THREE 30 mm BARRELS,** Vuk Garčević, Andrija Veljović, Igor Radisavljević
10.5937/OTEH26019G

114 **PERFORMANCE ANALYSIS OF SPECIAL SUBSONIC SMALL CALIBER WEAPONS FOR URBAN CONDITIONS,** Nebojša Hristov, Damir Jerković, Aleksa Aničić, Miljko Pavlović, Igor Epler
10.5937/OTEH26020H

121 **PRACTICAL APPROACHES TO QUALITY TESTING OF PIONEER MACHINES,** Dragan Stojadinović, Aleksandar Đurić, Snežana Jovanović, Miroslav Jovanović, Sreten Perić
10.5937/OTEH26021S

127 **IMPLEMENTATION AND EXPERIMENTAL VERIFICATION OF ENVIRONMENTAL SAFETY SENSOR DEVICE FOR FUZES IN AERIAL BOMBS,** Slobodan Pelemiš, Nikola Latinović, Mladen Veinović
10.5937/OTEH26022P

132 **MONTE CARLO BASED PRECISION ANALYSIS OF A LONG-RANGE ARTILLERY PROJECTILE,** Damir D. Jerković, Nebojša Hristov, Aleksandra Obradović, Nikola Mačar, Branislav Nikolić
10.5937/OTEH26023J

138 **SHEAR STRESS-BASED FORCE MODELING FOR M-84 TRACKED VEHICLE ON DEFORMABLE TERRAIN,** Momir Drakulić, Aleksandar Đurić
10.5937/OTEH26024D

144 **A NEW METHOD FOR INTERPOLATION OF ROCKET MOTOR THRUST - TIME CURVE FOR ARBITRARY PROPELLANT TEMPERATURE,** Miodrag Ćurčin, Goran Marjanović, Srđan Stojković, Bojan Tanaskovski
10.5937/OTEH26025C

152 **ANALYSIS OF RECOIL RESISTANCE FORCE BEHAVIOR AS A FUNCTION OF VARIATIONS IN INFLUENTIAL PARAMETERS,** Milan Simonović, Marko Simonović, Nikola Radovanović, Andrija Veljović, Ana Simonović, Sanja Milinković
10.5937/OTEH26026S

159 **DYNAMIC ANALYSIS OF 30 mm AUTOMATIC CANNON OPERATING MECHANISM,** Nikola Radovanović, Aleksandar Kari, Milan Simonović, Marko Simonović
10.5937/OTEH26027R

165 **INTERNAL BALLISTIC DESIGN OF A 30 x 173 mm AUTOMATIC CANNON USING THE ANALYTIC METHOD,** Marija Milovanović, Nikola Radovanović, Stefan Jovanović
10.5937/OTEH26028M

173 **TOOTH ROOT LOAD-CARRYING CAPACITY OF SPUR GEARS UNDER STANDARD SINGLE AND DOUBLE MESH CONDITIONS,** Vuk Milanović, Predrag Dobratić, Miloš Sedak, Dragan Trifković
10.5937/OTEH26029M

179 **SURROGATE-ASSISTED OPTIMIZATION OF PLANETARY GEARBOXES USING GENETIC ALGORITHM AND DEEP NEURAL NETWORK EFFICIENCY MODELS,** Miloš Sedak, Maja Rosić
10.5937/OTEH26030R

185 **DESIGN, MANUFACTURING AND TESTING OF A FUNCTIONAL MODEL OF A DEVICE WITH A SECURING EXECUTIVE MECHANISM FOR MULTIPLE REMOTE ACTIVATION OF PYROTECHNIC ELEMENTS ON UNMANNED PLATFORMS,** Slobodan Bosiljčić, Miroslav Gluvačević, Goran Marjanović, Jelena Petković Cvetković, Bojan Tanaskovski
10.5937/OTEH26031B

190 **ANALYSIS OF OSCILLATIONS OF A MOBILE PLATFORM UNDER THE ACTION OF FIRING FORCE,** Marko Simonović, Milan Simonović, Pera Srbul, Ana Simonović, Vuk Garčević
10.5937/OTEH26032S

197 **CHALLENGES OF DRONE-DELIVERED EXPLOSIVE ORDNANCE:TECHNICAL, OPERATIONAL, AND LEGAL-ETHICAL CONSIDERATIONS,** Vinko Žnidaršič, Stevan M. Mušicki, Nenad V. Kovačević
10.5937/OTEH26033Z

201 **BIASED PURSUIT GUIDANCE LAW FOR REAR-SECTOR TARGET INTERCEPTION,** Milos Pavić, Saša Antonović, Ivan Marić
10.5937/OTEH26034P

204 **CALIBRATION OF DIAL INDICATORS,** Dragan Lazić, Žaklina Čutović, Miloš Jovanović
10.5937/OTEH26035L

209 **TOTAL QUALITY INTEGRATED IN NEW AMMUNITION PRODUCTS IN AD "PRVI PARTIZAN" (ACCORDING TO ISO/TR 14062 - 2002),** Svetolik Spasenić
10.5937/OTEH26036S

213 **OPTIMIZATION OF CHASSIS GEOMETRY FOR INCREASING THE PRODUCTION CAPACITY OF 4X4 VEHICLE PLATFORM,** Mihajlo Vasiljević, Bogdan Živković, Jovanka Lukić
10.5937/OTEH26037V

220 **INTERNAL BALLISTICS-BASED DESIGN AND STRUCTURAL ANALYSIS OF A 30×165 mm CANNON BARREL,** Mihailo Bojković, Vladimir Dunić, Vladimir Milovanović
10.5937/OTEH26038B

226 **CONCEPTUAL ANALYSIS OF A RADIAL HYDRAULIC MOTOR DRIVE FOR HEAVY DUTY MILITARY VEHICLES,** Mihailo Đuković, Saša Milojević, Slobodan Savić, Milan Bukvić, Horia Abaitancei, Blaža Stojanović
10.5937/OTEH26039D

233 **PREDICTION ALGORITHM FOR THE IMPACT POINT PROJECTILE TRAJECTORY USED FOR DETERMINING CORRECTIVE SHOOTING PROCEDURE ELEMENTS,** Dragoslav R. Petrović
10.5937/OTEH26040P

4. SECTION : AMMUNITION AND ENERGETIC MATERIALS - AEM

240 **STABILITY EVALUATION OF MULTI-COMPONENT PBX UNDER VARYING CLIMATIC CONDITIONS,** Mirjana Krstović, Teodora Stančić, Vesna Petrović, Katarina Nestorović, Jelena Šultans, Danica Bajić
10.5937/OTEH26041K

247 **EXPERIMENTAL AND NUMERICAL STUDY OF PRECURSOR SHAPED CHARGE PERFORMANCE,** Predrag Elek, Biljana Jovanović, Dragana Jelić, Danijela Obradović, Danica Jelić, Dušan Blagojević
10.5937/OTEH26042E

254 **ANALYSIS OF ENERGY DISSIPATION DURING THE IMPACT OF AIR WEAPON PROJECTILES INTO GRANULAR MATERIALS – COMPARISON OF ANALYTICAL MODEL AND EXPERIMENTAL RESULTS,** Darko Jankovic, Marina Simovic Pavlovic, Nada Citakovic, Lana Milosavljevic, Ljubisa Tomic, Martin Ocsenhofer, Darko Vasiljevic
10.5937/OTEH26043J

259 **SAPONIFICATION AS AN ENABLING STEP FOR FAST GC DETERMINATION OF STABILIZERS IN SMOKELESS POWDERS,** Teodora Stančić, Nada Marinković, Suzana Kovačević, Mirjana Milekić, Vesna Petrović, Mirjana Krstović
10.5937/OTEH26044S

265 **ANALYTICAL DETERMINATION OF MEAN BALLISTIC PRESSURE FROM THE ENERGY BALANCE OF PROPELLANT COMBUSTION PRODUCTS WITH EXPERIMENTAL VALIDATION,** Slobodan Orašanin, Miodrag Lisov, Aleksa Aničić, Miloš Jončić, Miljko Pavlović
10.5937/OTEH26045O

271 **DEFINITION OF A METHOD FOR VERIFYING THE QUALITATIVE CHARACTERISTICS OF ELECTRIC IGNITER HEADS,** Jelena Petković-Cvetković, Branka Džunić, Slobodan Bosiljčić, Goran Marjanović, Dragan Knežević, Jelena Mojsilović, Marko Radisavljević
10.5937/OTEH26046P

277 **MODIFIED PBXN-5 EXPLOSIVE: EFFECT OF IF-WS2 NANOPARTICLES ON PHLEGMATIZATION OF HMX,** Danica Bajić, Mirjana Krstović, Ivan Dimitrijević, Mladen Timotijević, Ognjen Rajković, Jelena Marinković
10.5937/OTEH26047B

283 **APPLICATION OF A CASCADE OF NEURAL NETWORKS FOR MODELING AND PREDICTING DETONATION CHARACTERISTICS OF HIGH-ENERGY MATERIALS,** V.S. Abrukov, D.A. Anufrieva, S.S. Kuzmin, D.A. Saperov, I.N. Ulyanov
10.5937/OTEH26048A

287 **NUMERICAL–EXPERIMENTAL ANALYSIS OF EXPLOSIVELY FORMED PROJECTILE PENETRATION CHARACTERISTICS,** Andjela Ivković, Sasa Savić, Jelena Marinković, Mladen Timotijević
10.5937/OTEH26049I

292 **DETERMINING LAW COMBUSTION PARAMETERS AND THEIR PRESSURE/COMPOSITION DEPENDENCES FOR HIGH-ENERGY MATERIALS VIA NEURAL NETWORK SEQUENTIAL APPROXIMATIONS,** D.A. Anufrieva, V.S. Abrukov, Vladica Bozic, D.A. Sapyerov, I.N. Ulyanov
10.5937/OTEH26050A

296 **INFLUENCE OF GROOVE GEOMETRY ON THE INTERNAL BALLISTICS OF A MONOLITHIC .308 CALIBER BULLET,** Igor Babić, Mihailo Erčević
10.5937/OTEH26160B

303 **TBX FORMULATIONS WITH ALTERNATIVE METAL FUELS, HIGH EXPLOSIVES AND OXIDIZERS,** Thomas M. Klapötke, Danica Bajić, Stanisław Cudziło, Muhamed Sućeska
10.5937/OTEH26051K

5. SECTION : INTEGRATED SENSOR SYSTEMS AND ROBOTIC SYSTEMS – ISSRS

308 **SIMULATION OF COMPLIANT ROBOTIC SYSTEM WITH A PARALLEL MECHANISM AND SCISSOR LIFT**, Nemanja Tanasković, Damir Krklješ, Mihailo Lazarević
10.5937/OTEH26052T

315 **REAL-TIME THERMAL IMAGE SYNTHESIS IN UNITY 3D VIA EMPIRICAL INFRARED ADAPTATION OF LIGHT SIMULATION MODELS**, Bojan Šmidling, Stefan Stojkov
10.5937/OTEH26053S

321 **IMPROVING DRONE DETECTION USING ACOUSTIC METHODS AND STATISTICAL FILTERING,** Miodrag Vračar, Stevo Vračar, Radiana Brusin
10.5937/OTEH26054V

327 **ADVANCED HYDROGEN DETECTION MATERIALS FOR CBRN APPLICATIONS: FROM METAL–ORGANIC FRAMEWORKS TO REAL-TIME THREAT ASSESSMENT,** Ivana Perović, Snežana Brković, Mina Seović, Stefan Mitrović, Milena Pijović Radovanović, Jelena Georgijević, Petar Laušević
10.5937/OTEH26055P

333 **INTERACTING MULTIPLE MODEL–BASED ESTIMATION OF AIRSPEED AND ALTITUDE FROM PITOT–STATIC MEASUREMENTS,** Vlada Sokolović, Momir Stanković, Goran Marković

10.5937/OTEH26056S

339 **SENSOR FUSION FOR DRONE CLASSIFICATION IN RADIO FREQUENCY DOMAIN,** Darko Mikanović, Boban Sazdić-Jotić, Vasilija Joksimović

10.5937/OTEH26057M

345 **FEDERATED LEARNING FOR AUTONOMOUS DISTRIBUTED GNSS JAMMING SWARMING STATIONS,** Boban Sazdić-Jotić, Dragana Bojić, Radiana Brusin

10.5937/OTEH26058S

352 **A COMPARATIVE EVALUATION OF SPECTROGRAM- AND BISPECTRUM-BASED TRANSFER LEARNING FOR DRONE AUDIO CLASSIFICATION,** Sara Milanović, Miodrag Vračar, Ivan Pokrajac

10.5937/OTEH26059M

358 **IMPACT OF JPEG COMPRESSION ON YOLO-BASED ROAD SCENE OBJECT DETECTION IN VISIBLE AND THERMAL IMAGES,** Boban Pavlović, Boban Bondžulić, Dimitrije Bujaković, Momir Stanković

10.5937/OTEH26060P

365 **AUTONOMOUS ROBOTIC SYSTEMS FOR 3D MOBILE MAPPING IN UNDERGROUND MINING,** Nenad Vušović, Milica Vlahović

10.5937/OTEH26061V

371 **EDGE AI DRONE CLASSIFICATION IN RADIO FREQUENCY DOMAIN,** Vasilija Joksimović, Darko Mikanović, Radiana Brusin, Dragana Bojić

10.5937/OTEH26062J

377 **METHODS FOR IMPROVING MACHINE LEARNING MODELS FOR PREDICTING MULTI-HOLE PRESSURE PROBE CHARACTERISTICS,** Ivana Petković, Snežana Zurovac, Dijana Damljanović, Ilija Nenadić

10.5937/OTEH26063P

382 **PERFORMANCE EVALUATION OF PSO TUNED FRACTIONAL ORDER ILC FOR THE LOLA15 ROBOT,** Nikola Živković, Mihailo Lazarević, Jelena Vidaković

10.5937/OTEH26064Z

388 **METHOD AND SYSTEM FOR THROUGH-THE-WALL RADAR IMAGING,** Mladen Živković

10.5937/OTEH26065Z

393 **AUTOMATED IMAGE CAPTIONING WITH A RESNET-LSTM ENCODER-DECODER FOR DISTRIBUTED SURVEILLANCE SCENES,** Anđa Denić, Snežana Zurovac, Ivana Petković

10.5937/OTEH26066D

401 **SWARM INTELLIGENCE: DIFFERENCES BETWEEN BIOLOGICAL PRINCIPLES AND ARTIFICIAL INTELLIGENCE CAPABILITIES,** Miloš Jovanović, Jelena Vasiljević, Igor Medenica, Dragan Lazić, Calin-Octavian Micloshina

10.5937/OTEH26067J

407 **APPLICATION OF THE BARTLETT METHOD FOR TARGETS ANGLE ESTIMATION IN FMCW RADAR SYSTEMS,** Borko Đaković, Slobodan Simić, Nikola Danilović

10.5937/OTEH26068D

413 **COMPARATIVE ANALYSIS OF YOLO MODELS FOR OBJECT DETECTION IN IMAGES COMPRESSED WITH JPEG, BPG, AND VVC,** Nikola Bukvić, Boban Bondžulić, Dimitrije Bujaković, Milenko Andrić, Boban Pavlović
10.5937/OTEH26069B

420 **STRUCTURAL INTEGRITY AND FEM-BASED OPTIMIZATION OF A MISSILE FIN SERVO MECHANISM,** Nebojša Jovičić, Marko Bojanić, Aleksandar Stefanović, Marijana Stojanović
10.5937/OTEH26070J

425 **INVERSE THERMOVISION ANALYSIS OF HIGH-ENERGY PHENOMENA,** Dragan M. Knežević, Milan Popović, Katarina Mišković, Milica Marjanović, Jelena Petković-Cvetković
10.5937/OTEH26071K

431 **REACH AND WORKSPACE ENVELOPE ANALYSIS OF THE SIX-DOF ARTICULATED ROBOT ARMA6,** Jelena Vidaković, Andrija Dević, Goran Vasilić, Nikola Živković, Ivana Vasović Maksimović
10.5937/OTEH26072V

437 **STOCHASTIC ERROR CHARACTERIZATION AND KALMAN FILTERING OF A MEMS GYROSCOPE BASED ON ALLAN VARRIANCE ANALYSIS,** Adrijana Jovanović, Jelena Popović-Božović, Zoran Krstić
10.5937/OTEH26073J

444 **A LAYERED COUNTER-UAS CONCEPT: FROM MULTI-SENSOR DETECTION AND CLASSIFICATION TO KILL-WEB INTEGRATION,** Ilija Popadić, Miljan Vučetić, Zoran Banjac, Boško Božilović, Ivan Vulić
10.5937/OTEH26074P

450 **AUTONOMOUS LOCOMOTION OF ROBOTIC DOGS IN MILITARY APPLICATIONS,** Nina Mitričević, Rade Pavlović
10.5937/OTEH26075M

455 **UAV POSITION ESTIMATION THROUGH IMAGE MATCHING BETWEEN DRONE IMAGERY AND MAPS,** Rade Pavlović, Sara Pavlović
10.5937/OTEH26076

460 **GENETIC ALGORITHM PARAMETER CONTROL FOR THE OPTIMAL DESIGN OF SPECTRALLY SELECTIVE 1D PHOTONIC CRYSTALS,** Anja Perić, Olga Jakšić
10.5937/OTEH26139P

6. SECTION : TELECOMMUNICATION AND INFORMATION SYSTEMS - TIS

465 **AN EXAMPLE OF MISSILE-TO-TARGET INTERCEPTION,** Zvonko Radosavljević, Dragan Nikolić, Darko Pijević, Dejan Ivković
10.5937/OTEH26077R

470 **DEEPFAKE TECHNOLOGY AND THE TRANSFORMATION OF TRUST IN CONTEMPORARY INFORMATION SYSTEMS,** Sanela Veljković, Marina Dabetić, Anita Klikovac
10.5937/OTEH26078V

475 **A MULTI-OBJECTIVE APPROACH TO OPTIMIZING VIRTUALIZED MILITARY DATA CENTER INFRASTRUCTURE WITH GENETIC ALGORITHMS,** Nikola Zogović, Sava Stanišić, Borislav Đorđević
10.5937/OTEH26079Z

482 **A MULTI-CRITERIA APPROACH FOR AUTOMATION OF REAL-TIME WEAPON ASSIGNMENT PROCESS,** Darko Pijević, Ksenija Bokić, Ina Masnikosa, Tanja Lacmanović

10.5937/OTEH26080P

488 **SOFTWARE-DEFINED RADAR DIGITAL EXCITER, RECEIVER, AND SIGNAL PROCESSOR,** Miloš D. Jevtić, Stefan Stojkov, Nikola Luković, Miloš Otović, Srđan Radovanović, Ina Masnikosa, Dragan Nikolić, Darko Pijević, Dejan Ivković, Zvonko Radosavljević

10.5937/OTEH26082J

494 **PREVENTIVE EM MEASUREMENTS: COST-EFFECTIVE SOLUTION FOR AWARENESS OF EMERGING TEMPEST THREATS,** Dragan Kondić, Borko Djaković, Milena Grdović, Nenad Stefanović

10.5937/OTEH26081K

502 **MODEL-BASED DESIGN AND IMPLEMENTATION OF AN OFDM TRANSMITTER FOR STM32F407 VIA MATLAB/SIMULINK,** Saša Gundelj, Pavel Crnomarković, Srđan Trandafilović, Igor Goršić

10.5937/OTEH26083G

507 **TELECOMMUNICATION AND INFORMATION SYSTEMS INFORMATION-PSYCHOLOGICAL ACTIVITIES IN MODERN CONFLICTS,** Jelena Bajić

10.5937/OTEH26084B

512 **UAV DETECTION AND TRACKING IN VIDEO USING DEEP LEARNING,** Lazar Velinov, Jovan Radivojević

10.5937/OTEH26085V

517 **EXTENDING THE XTTS MODEL FOR SERBIAN TEXT-TO-SPEECH GENERATION,** Stefan Ivanović, Dimitrije Kolašinac, Jordan Atanasijević, Kristina Živanović

10.5937/OTEH26086I

523 **PASSIVE TIME OF ARRIVAL BASED TARGET LOCALIZATION FOR RADAR SYSTEMS USING HYBRID METAHEURISTIC OPTIMIZATION ALGORITHMS,** Maja Rosić, Miloš Sedak

10.5937/OTEH26087R

529 **EFFICIENCY ANALYSIS OF A STEGANOGRAPHY METHOD BASED ON PIXEL VALUE DIFFERENCES,** Sanja Klajić, Branko Vujatović, Darko Grubač, Nenad Stojanović

10.5937/OTEH26088K

535 **EVALUATION OF TRANSFORMER-BASED EMBEDDING MODELS FOR INFORMATION RETRIEVAL IN SERBIAN MILITARY TECHNICAL TEXTS,** Slavimir Stošović, Nikola Vukotić

10.5937/OTEH26089S

540 **A RETRIEVAL-AUGMENTED GENERATION SYSTEM FOR SERBIAN MILITARY TECHNICAL DOCUMENTATION USING LOCAL LANGUAGE MODELS,** Nikola Vukotić, Slavimir Stošović

10.5937/OTEH26090V

545 **DEEPFAKE DETECTION IN DEFENSE COMMUNICATION SYSTEMS: EVALUATION OF REAL-TIME DETECTION APPROACHES,** Milan Ristić, Nikola Vukotić

10.5937/OTEH26091R

551 **EVALUATING THE EFFECTIVENESS OF THE LOCALLY DEPLOYED GPT-OSS MODEL IN DETECTING SMB AND RDP BRUTE-FORCE ATTACKS FROM WINDOWS SECURITY LOGS,** Igor Petrović, Milomir Jovićević, Marko Premović, Mladen Veinović, Ivan Tot

10.5937/OTEH26092P

556 **POSITION BIAS IN MULTILINGUAL LANGUAGE MODELS APPLIED TO SERBIAN,** Milomir Jovićević, Slaviša Ilić, Milan Gnjatović, Kristina Živanović

10.5937/OTEH26093J

562 **INTEGRATING ARTIFICIAL INTELLIGENCE AND REVERSE ENGINEERING INTO CONSTRUCTIVE LEARNING FOR CYBERSECURITY EDUCATION IN INDUSTRY 4.0,** Edis Mekić, Ana Matović, Marija Matović

10.5937/OTEH26094M

568 **MRC COMBINER WITH THREE INPUTS IN THE PRESENCE OF FAST RICE FADING,** Ana Matović, Marija Matović, Elmedin Biberović

10.5937/OTEH26095M

573 **QUANTITATIVE ASSESSMENT OF ZERO TRUST ARCHITECTURE EFFECTIVENESS AGAINST ADVANCED CYBER ATTACKS IN A SMALL-SCALE EMULATED GOVERNMENT AND MILITARY INFORMATION SYSTEMS,** Boban Stojanović, Mladen Veinović, Marko Premović, Igor Petrović

10.5937/OTEH26096S

579 **CHARACTERIZATION AND SUPPRESSION OF COMMON-MODE AND DIFFERENTIAL-MODE CONDUCTED EMI,** Pavel Crnomarković, Saša Gundelj, Mihailo Knežević, Radoslav Surla

10.5937/OTEH26097C

585 **GEOSPATIAL ASSESSMENT OF THE DEVELOPMENT LEVEL OF LOCAL GOVERNMENT UNITS IN THE REPUBLIC OF SERBIA,** Zoran Kričković, Tin Lukić, Saša Bakrač, Vladan Tadić, Emina Kričković

10.5937/OTEH26098K

591 **AUTOMATING DIGITAL MAP PRODUCTION PROCESSES USING PYTHON GIS ALGORITHMS,** Marko Simić, Saša Bakrač, Đorđe Đermanović, Matija Milić, Slaviša Tatomirović

10.5937/OTEH26099S

598 **COMPARATIVE EVALUATION OF SIEM PLATFORMS IN DETECTING MITRE ATT&CK TECHNIQUES BASED ON FALSE POSITIVE ANALYSIS,** Marko Premović, Mladen Veinović, Boban Stojanović, Igor Petrović

10.5937/OTEH26100P

605 **MODULARITY OR COMPRESSION? A UNIFIED COMPARISON OF MIXTURE-OF-EXPERTS AND POLICY DISTILLATION AS EXECUTION MODULES FOR LLM-PLANNED AGENTS,** Igor Ciganović, Aleksandar Pluškoski, Miloš Jovanović, Jelena Vasiljević

10.5937/OTEH26101C

612 **AI-ASSISTED DECISION SUPPORT IN COMMAND-AND-CONTROL SYSTEMS: APPLYING THE SENSE–UNDERSTAND–DECIDE–ACT (SUDA) CYCLE,** Ivan Vulić, Miljan Vučetić, Ilija Popadić, Ranko Petrović, Zoran Banjac

10.5937/OTEH26102V

7. SECTION : MATERIALS AND TECHNOLOGIES - MT

620 **SOLITON-MEDIATED ENERGY TRANSFER AND MAGNETIZATION CONTROL IN PLD-GROWN E-FE_2O_3 THIN FILMS,** Violeta N. Nikolić, Jose F. M. L. Mariano
10.5937/OTEH26103N

626 **FLOQUET-ENHANCED ELECTRON TRANSPORT IN 1D Fe–X MAGNETIC MOLECULAR SPINTRONIC CHAINS,** Jose F. M. L. Mariano, Violeta N. Nikolić
10.5937/OTEH26104M

633 **THERMAL BEHAVIOR OF THE HIERARCHICAL STRUCTURE OF MORPHO DIDIUS BUTTERFLY WINGS: IMPLICATIONS FOR PASSIVE INFRARED SIGNATURE MANAGEMENT,** Marina Simović – Pavlović, Luka Marjanović, Katarina Nestorović, Darko Vasiljević
10.5937/OTEH26105S

639 **EXPERIMENTAL EVALUATION OF THE TENSILE PROPERTIES OF DIFFERENT TYPES OF ARMOUR STEEL AND THEIR WELDED JOINTS,** Vladimir Milovanović, Dušan Arsić, Dušan Popadić, Miloš Pešić, Katarina Dragutinović
10.5937/OTEH26106M

645 **FLEXIBLE ELECTROCHEMICAL ENERGY SOURCES FOR WEARABLE SYSTEMS: TECHNICAL SOLUTIONS AND INTEGRATION WITH A MILITARY APPLICATION ASPECT,** Božana Gaćić, Mihael Bučko, Milica Košević, Radovan Karkalić, Sreten Perić
10.5937/OTEH26107G

652 **PERFORMANCE AND DEVELOPMENT TRENDS OF PERSONAL BALLISTIC PROTECTIVE EQUIPMENT IN THE CONTEXT OF MODERN THREATS,** Siniša Gaćić, Nikola Prodanović, Marko Milovanović
10.5937/OTEH26108G

661 **CARBON-BASED ADSORPTIVE MATERIALS FOR MITIGATION OF ORGANOPHOSPHATE CONTAMINANTS IN AQUATIC SYSTEMS,** Vladan J. Anićijević, Katarina Kokanov Stanković, Vedran Milanković, Tamara Terzić, Radovan M. Karkalić, Tamara Lazarević-Pašti
10.5937/OTEH26109A

666 **SYNTHESIS OF TITANIUM BASED POWDERS,** Srećko Stopić, Bernd Friedrich, Duško Kostić; Mitar Perušić, Vladimir Damjanović, Radislav Filipović, Mirko Stamatović, Dejan Jovanović
10.5937/OTEH26110S

670 **POROUS CERAMIC FOAM AS A MEDIUM FOR COBALT IMMOBILIZATION IN RADIONUCLIDE WATER TREATMENT,** Denis Dinić, Milan Tanić, Jelena Dikić, Stevan Stupar, Nemanja Jovanović, Sanja Jevtić
10.5937/OTEH26111D

675 **INFLUENCE OF SURFACE PREPARATION ON THE MECHANICAL PROPERTIES OF ADHESIVELY BONDED JOINTS IN METAL STRUCTURES,** Nataša Zdravković, Dragan Milčić, Damjan Klobčar, Tomaž Vuherer, Miodrag Milčić, Vukašin Pavlović, Aleksija Đurić
10.5937/OTEH26112Z

679 **BEHAVIOR OF FLAT SAMPLES MADE OF S355 STRUCTURAL STEEL UNDER STATIC LOADING AND STRESS CONCENTRATION,** Jovana Antić, Mileta Ristivojević, Aleksandar Dimić, Stefan Dikić, Ana Petrović, Saša Spasenić, Sreten Spasenić
10.5937/OTEH26113A

684 **TEMPORAL VARIATIONS IN BACKGROUND RADIATION AND THEIR CORRELATION WITH METEOROLOGICAL PARAMETERS IN VRANJE, SERBIA,** Milan Tanić, Denis Dinić, Mladen Nikolić, Stevan Stupar
10.5937/OTEH26114T

691 **INVESTIGATION OF NON-STOICHIOMETRIC TUNGSTEN OXIDES AS SUPPORTS FOR ANODE CATALYSTS IN PEM FUEL CELLS,** Snežana Brković, Ivana Perović, Mina Seović, Stefan Mitrović, Milena Pijović Radovanović, Jelena Georgijević, Petar Laušević
10.5937/OTEH26115B

695 **EVALUATION OF B-PINENE AS AN ENVIRONMENTALLY FRIENDLY CORROSION INHIBITOR FOR CARBON STEEL IN 1 M HYDROCHLORIC ACID,** Anđela Simović, Jelena Bajat
10.5937/OTEH26116S

700 **FRICTION STIR WELDING IN THE JOINING OF LIGHT METAL ALLOYS: A REVIEW,** Nada Ratković, Vesna Mandić, Živana Jovanović Pešić, Marko Delić, Đorđe Ivković
10.5937/OTEH26117R

707 **GRAPHENE OXIDE FOR ORGANOPHOSPHATE REMOVAL – A DFT INSIGHT,** Dragana Vasić Anićijević, Vladan Anićijević
10.5937/OTEH26118V

711 **ADDITIVE MANUFACTURING IN THE MILITARY INDUSTRY: RESILIENCE AND SUSTAINABILITY FOR SMALL STATES,** Saša Veselinović, Stevan Jovičić, Dragan Stojadinović
10.5937/OTEH26119V

717 **INTEGRATED ELECTROCHEMICAL ASSESSMENT OF STEEL CORROSION IN CHLORIDE ENVIRONMENTS,** Milica Vlahović, Marina Dojčinović, Đorđe Gjumišev, Nataša Đorđević, Jovanka Pejić
10.5937/OTEH26120V

722 **APPLICATION OF XRD AND AAS FOR EVALUATION OF MECHANOCHEMICAL REACTION PROGRESS,** Nataša Đorđević, Srđan Matijašević, Nenad Vušović, Mladen Pejić, Slavica Mihajlović, Milica Vlahović
10.5937/OTEH26121D

727 **COMPARATIVE ANALYSIS OF DIFFERENT 3D PRINTED IGNITER CHARGES ON INTERIOR BALLISTIC PROPERTIES,** Veselin Živanović, Jelena Mojsilović, Marko Radisavljević, Vera Joksimović, Jovana Mandić, Irina Ulanova
10.5937/OTEH26122Z

731 **THE EFFECT OF PALLADIUM LOADING ON PD-MODIFIED SB CATALYSTS IN THE ELECTROOXIDATION OF ETHANOL,** Milica G. Košević, Nebojša D. Nikolić, Jelena D. Lović
10.5937/OTEH26123K

735 **BULLETPROOF/BULLET RESISTANT CERAMIC MATERIALS FOR ARMOR APPLICATIONS,** Tatjana Volkov-Husović, Ana Alil, Sanja Martinović
10.5937/OTEH26124V

740 **ADSORPTION OF AN ORGANIC POLLUTANT ONTO BIOCHAR DERIVED FROM SOUR CHERRY STONE,** Anja Antanasković, Milan Milivojević, Nikola Vuković, Vladimir Adamović, Tatjana Šoštarić, Zorica Lopičić
10.5937/OTEH26125A

744 **INFLUENCE OF PARTICLE SIZE ON ADSORPTION PROPERTIES OF GROUND FOAM GLASS,** Veljko Savić, Anja Antanasković, Tamara Vujatović-Velimirov, Vladimir Topalović, Zorica Lopičić, Jelena Nikolić, Snežana Grujić

10.5937/OTEH26126S

748 **RADIOACTIVITY IN FOOD SAMPLES MEASURED IN THE RADIOCHEMICAL LABORATORY OF THE MILITARY TECHNICAL INSTITUTE,** Nataša Pajić, Anton Radojković

10.5937/OTEH26127P

755 **INFLUENCE OF THERMOMECHANICAL TREATMENT ON THE GRAIN STRUCTURE OF Al-Mg-Mn ALLOYS FOR SHEAR SPINNING APPLICATION,** Ljubica Radović, Miljana Popović, Branka Jordović

10.5937/OTEH26128R

761 **DIFFERENT SYNTHETIC ROUTES OF CERIUM-DIOXIDE NANOPARTICLES PREPARATION AND THEIR EFFICIENCY IN SULFUR MUSTARD REMOVAL,** Maja Vitorović-Todorović, Aleksandar Ćitić

10.5937/OTEH26129V

767 **INFLUENCE OF AN INCREASED HOPCALITE LAYER CONTENT IN THE ACTIVE FILLING ON THE SORPTION CHARACTERISTICS OF THE FILTER,** Biljana Mihajlović, Tatjana Marković, Marina Ilić, Vukica Grković, Vlada Petrovic

10.5937/OTEH26130M

773 **INFLUENCE OF PROCESS PARAMETERS ON QUALITY OF LASER POWDER BED FUSION PRODUCED AL10SIMG PARTS,** Aleksandar Ćitić, Vesna Lukić, Jovana Mandić, Marko Bojin, Aleksa Grubić, Miloš Stepanović, Živorad Jevtić

10.5937/OTEH26131C

780 **EFFECT OF SIO2 NANOPARTICLES ON THE MECHANICAL PROPERTIES OF POLYURETHANE THERMAL INSULATION LINERS,** Sara Pepić, Jelena Gržetić, Olga Maraš, Veselin Živanović, Teodora Stančić

10.5937/OTEH26132P

786 **NUMERICAL ANALYSIS OF THE HOT FORGING PROCESS FOR A STEEL DEFENSE INDUSTRY COMPONENT**, Nikola Kastratović, Vesna Mandić, Marko Delić, Dušan Arsić, Nikola Komatina

10.5937/OTEH26133K

792 **CAE SIMULATION OF THE HOT FORGING PROCESS FOR A MILITARY VEHICLE STEEL COMPONENT**, Marko Delić, Nikola Kastratović, Vesna Mandić, Dušan Arsić, Nikola Komatina

10.5937/OTEH26134D

798 **OPTIMIZATION OF INTERFACIAL BONDING OF EPDM-BASED THERMAL INSULATION AND STEEL ROCKET MOTOR CASE-DEVELOPMENT OF THE SYSTEM OF ADHESIVE LAYER AND CEMENT,** Olga Maraš, Sara Pepić, Jelena Gržetić, Petar Vasiljević

10.5937/OTEH26135M

804 **A COMPARATIVE ASSESSMENT OF DIFFERENT NERVE AGENT VX STOCKS USING CHOLINESTERASE INHIBITION ASSAYS,** Sonja Bauk, Maja Vitorović-Todorović, Marina Ilić

10.5937/OTEH26136B

810 **MAGNETOIMPEDANCE EFFECT OF $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ AMORPHOUS RIBBON,** Radoslav Surla, Jovana Mandić, Jelena Orelj, Nebojša Mitrović, Pavel Crnomarković
10.5937/OTEH26137S

815 **CORRELATION BETWEEN IMPACT ENERGY AND THICKNESS IN QUENCHED AND TEMPERED LOW ALLOYED STEEL,** Miloš Čenić, Esmail Ali Salem Ahmed, Srdja Perković, Nenad Radović, Dragomir Glišić, Stefan Dikić
10.5937/OTEH26138C

8. SECTION: QUALITY, STANDARDIZATION, METROLOGY, MAINTENANCE AND EXPLOITATION - QSMME

820 **FINANCIAL SUSTAINABILITY OF THE DEFENSE SYSTEM OF THE REPUBLIC OF SERBIA: BUDGETARY AND ACCOUNTING ASPECTS,** Aleksandar Savić, Milan Mihajlović, Aleksandar Rajković
10.5937/OTEH26140S

827 **VARIABILITY OF BALLISTIC TEST RESULTS AND THE RELIABILITY OF QUALITY ASSESSMENT OF PERSONAL BALLISTIC PROTECTION EQUIPMENT,** Marko Milovanović, Siniša Gaćić, Nikola Prodanović
10.5937/OTEH26141M

833 **INSPECTION OF THE PROPULSION SYSTEM OF THEAUXILIARY SHIP PBR-36 "ŠABAC",** Enis Dauti, Dragan Trifković, Dušan Cincar
10.5937/OTEH26142D

839 **MECHANICAL QUALIFICATION OF SINGLE SOURCE COMMERCIAL OFF THE SHELF (COTS) ASSEMBLIES FOR AEROSPACE AND DEFENCE ENVIRONMENTS,** Konstantin Komatina
10.5937/OTEH26143K

844 **COMBINED APPLICATION OF FTA AND FMEA METHODS FOR FAILURE ANALYSIS OF TECHNICAL SYSTEMS,** Dobrivoje Ćatić, Vladimir Ćatić
10.5937/OTEH26144C

850 **A QUALITY ASSURANCE FRAMEWORK FOR METAL ADDITIVE MANUFACTURING,** Srdjan Živković, Slobodan Malbašić, Miloš Stepanović
10.5937/OTEH26145Z

857 **ELECTRO-MECHANICAL FUZES FOR UAV LOITERING MUNITION WARHEADS: A LAYERED SAFETY AND STANDARDIZATION FRAMEWORK,** Marinko Aleksić, Nikola Nikolić, Marija Borović, Dejan Marsenić, Danijela Kovač, Nemanja Vujović
10.5937/OTEH26146A

863 **CLIMATE CHANGES AND ITS IMPACTS ON WATER RESOURCES, CASE STUDY CRNICA RIVER BASIN,** Aleksandar Tanasković, Vesna Ristić Vakanjac, Veljko Marinović, Boris Vakanjac, Hristina Petrova
10.5937/OTEH26147T

871 **EVALUATING THE EFFECTIVE PERFORMANCE OF THE PHASE NOISE MEASUREMENT SYSTEM,** Neda Spasojević, Miša Markuš, Slobodan Subotić, Dragan Lazić
10.5937/OTEH26148S

875 **IMPLEMENTATION OF A LOW-COST HP-IB (GPIB, IEEE 488) INTERFACE AND MATLAB GUI APPLICATION FOR HP 3582A SPECTRUM ANALYZER CONTROL,** Petar Mišljen, Miloš Pavić, Saša Gundelj, Igor Goršić
10.5937/OTEH26149M

880 **QUASI-DYNAMIC TESTING OF THE QUALITY OF TRACER COMPOSITION IN SMALL ARMS AMMUNITION,** Boban Bogićević, Nikola Kilibarda, Miroslav Pajčin, Ognjen Peković, Aleksandar Simonović

10.5937/OTEH26150B

885 **RADIOACTIVE WASTE AS A LONG TERM SECURITY CHALLAENGE: PRINCIPLES OF PROTECTION AND RISK MANAGEMENT,** Milica Ćurčić, Predrag Božović, Marija Janković

10.5937/OTEH26151C

892 **EXPLOATATION OF ARMAMENT & MILITARY EQUIPMENT- INTERACTION MAN AND TECHNOLOGY: ORGANIZATIONAL CULTURE AS THE FOUNDATION OF MILITARY STANDARDS IN THE MODERN MILITARY PROFESSION,** Lazar Bjelica, Srđan Živković

10.5937/OTEH26152B

898 **THE INFLUENCE OF ULTRASONIC CAVITATION ON THE SURFACE ROUGHNESS OF LASER-SINTERED MS1 STEEL,** Pavle Ljubojević, Žarko Mišković, Tatjana Lazović

10.5937/OTEH26153L

902 **NATURAL AND ANTHROPOGENIC FACTORS INFLUENCING THE VELIKA MORAVA RIVER WATER REGIME,** Milan Kresojević, Vesna Ristić Vakanjac, Dušan Polomčić, Aleksandar Tanasković, Boris Vakanjac, Jugoslav Nikolic

10.5937/OTEH26154K

908 **COMPARATIVE ANALYSIS OF DIFFERENT SYSTEMS FOR PROJECTILE VELOCITY MEASUREMENT IN BALLISTIC TESTING,** Aleksa Aničić, Nebojša Hristov, Lazar Blagojević, Miloš Jončić, Ilija Ljujić, Damir Jerković

10.5937/OTEH26155A

915 **CFD-BASED NUMERICAL SIMULATION OF THERMAL PROCESSES AND AIRFLOW IN THE PASSENGER COMPARTMENT OF SPECIAL VEHICLE,** Milica Ivanović, Snežana Mališić, Marija Baltić, Igor Stamenković

10.5937/OTEH26156I

921 **ACCURACY ASSESSMENT OF GROUND TARGET COORDINATES DETERMINED BY DIFFERENT ACQUISITION SYSTEMS,** Srđan Stojaković Siniša Drobnjak, Novak Roganović, Saša Bakrač

10.5937/OTEH26157S

928 **NOVEL STRATEGIC TRENDS IN WEAPON SYSTEM ACQUISITION,** Nebojša Nikolić, Miloš Stojanović, Miroslav Mitrović

10.5937/OTEH26158N

933 **STANDARDIZATION AS A REGULATOR FOR THE USE OF ARTIFICIAL INTELLIGENCE IN FRAUD DETECTION,** Igor Đorić, Snežana Zurovac

10.5937/OTEH26159D

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION I

Aerodynamics and flight dynamics - AFD

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE STUDY OF HYBRID DOLPHIN AND OTHER BIONIC AIRFOILS FOR DEFENSE APPLICATIONS

ZORANA Z. DANČUO
University of Belgrade, Faculty of Mechanical Engineering, Innovation Center, Belgrade, zdancuo@mas.bg.ac.rs

IVAN A. KOSTIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, ikostic@mas.bg.ac.rs

OLIVERA P. KOSTIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, okostic@mas.bg.ac.rs

ALEKSANDAR Č. BENGIN
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, abengin@mas.bg.ac.rs

GORAN S. VOROTOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, gvorotovic@mas.bg.ac.rs

Abstract: *This paper presents a comparative study of hybrid dolphin and other bionic airfoils aimed at improving aerodynamic performance through a combined computational and experimental approach. Contemporary designs inspired by avian and marine species employ distinctive geometries and surface textures to enhance lift, reduce drag, and increase the lift-to-drag (L/D) ratio. A systematic comparative analysis of hybrid dolphin airfoils was conducted, evaluating their performance relative to established bionic optimization strategies reported in the literature. Aerodynamic characteristics were assessed using computational optimization techniques, including genetic algorithms and multi-island genetic algorithms, in conjunction with CFD simulations and experimental investigations across a range of Reynolds numbers. The obtained results demonstrate that optimized configurations consistently outperform baseline models, showing notable lift enhancement and drag reduction, leading to significant improvements in L/D ratio. These findings are consistent with literature reports indicating lift increases of up to 28.4%, drag reductions of up to 18.09%, and L/D enhancements of up to 323% under specific flow conditions. Such dramatic L/D improvements are associated with shark denticle-inspired surface structures at low angles of attack (below 4°) and Reynolds numbers of the order of* 4×10^4*. The comparative results confirm that bionic airfoil optimization provides measurable advantages over conventional designs, particularly within targeted operational regimes.*

Keywords: *optimisation, bionic airfoils, comparative analysis, aerodynamic efficiency, defense applications.*

1. INTRODUCTION

The continuous pursuit of improved aerodynamic efficiency is a key objective in the design of aerospace and defense systems, driven by the need to enhance performance, reduce fuel consumption, and expand operational capabilities. Among various optimization approaches, bio-inspired design has emerged as a promising strategy, drawing inspiration from the highly efficient aerodynamic and hydrodynamic adaptations observed in nature.

Bio-inspired design strategies in aerodynamics are based on the translation of morphological and functional characteristics observed in biological systems into engineering solutions for improved performance. In past decades, avian wing structures, marine animal skins, and aquatic locomotion mechanisms have served as primary sources of inspiration for enhancing aerodynamic efficiency in airfoil design [1, 2].

These strategies typically involve two directions. The first is geometric adaptation, in which the overall shape of the airfoil is modified to mimic profiles, such as those found in bird wings or dolphin-like shapes. The second direction focuses on surface modifications, including the incorporation of riblets, denticle-like textures, and micro coatings inspired by shark skin and other marine organisms, to reduce skin friction and minimize drag [3–7]. Along with morphological adaptations, contemporary and selected bionic optimization increasingly incorporates computational methods. Evolutionary algorithms, such as genetic algorithms and multi-island genetic algorithms, are commonly employed to explore large design spaces and identify configurations that yield optimal aerodynamic performance under given flow conditions [8, 9]. When combined with CFD simulations, these methods enable a detailed evaluation of complex bio-inspired geometries and their aerodynamic behaviour. Such approaches form an optimization framework in which biological principles, as well as numerical and experimental methods, are integrated. Despite significant advances in bio-inspired aerodynamic design, certain

Corresponding author: Z. Z. Dančuo;
E-mail: zdancuo@mas.bg.ac.rs
DOI: 10.5937/OTEH26001B

limitations remain in the current literature. Previous studies have demonstrated the potential of genetic algorithms and biomimetic surface structures to improve aerodynamic performance through lift enhancement and drag reduction [1-9]. However, most research focuses on optimization approaches, such as surface-texture adaptations. Furthermore, comparative analyses evaluating different bionic optimization strategies within a unified computational and experimental framework remain limited. In particular, the aerodynamic performance of hybrid dolphin-inspired airfoils has rarely been systematically compared with other bio-inspired optimization concepts under similar operating conditions.

To address this gap, this study provides a comparative assessment of hybrid dolphin airfoils [10-13] and selected bionic optimization strategies reported in the literature, using computational optimization, CFD simulations, and experimental investigations to evaluate their aerodynamic efficiency. The findings aim to support the development of more aerodynamically efficient configurations for aerospace and defense applications.

2. METHODS

2.1 Airfoil Selection

The Dolphin 0006 M4 airfoil (Figure 1) was selected from a seven-member hybrid Dolphin family due to its highest percentage improvement in lift coefficient. The study considers the hybrid Dolphin airfoils developed and analyzed by the authors [10–13], together with selected bio-inspired airfoil optimization concepts reported in the literature for subsequent comparative analysis. The analyzed configurations represent different bionic design approaches, including geometric modifications and biomimetic surface adaptations inspired by marine and/or avian species.

The selected studies were chosen because they reported quantitative aerodynamic improvements and represented distinct categories of bio-inspired aerodynamic optimization, including surface modifications, leading-edge modifications, and geometry-based airfoil designs.

2.2 Comparative Evaluation Procedure

A comparative assessment was performed using the obtained results for the hybrid Dolphin airfoils and the corresponding performance indicators reported in the literature. The analysis focused on lift enhancement, drag reduction, and L/D ratio improvements under the corresponding operating conditions.

2.3 Performance Metrics

The aerodynamic performance of the analyzed airfoils was evaluated using the lift coefficient (CL), drag coefficient (CD), and lift-to-drag ratio (L/D). These parameters were used to compare the effectiveness of different bionic optimization strategies and identify configurations offering improved aerodynamic efficiency.

2.4 Computational Fluid Dynamics Setup

Numerical simulations were performed to evaluate the aerodynamic characteristics of the investigated airfoil configurations under controlled flow conditions. The analysis was carried out using a finite volume-based CFD approach, solving the Reynolds-averaged Navier–Stokes (RANS) equations for incompressible flow. The computational domain was designed to capture wake development downstream of the airfoil accurately. A C-type structured mesh was employed, with local refinement in the vicinity of the airfoil surface and in the boundary layer region to properly resolve near-wall gradients. Special attention was given to mesh density in the leading- and trailing-edge regions, where strong pressure gradients and flow separation may occur [10-13].

Near-wall resolution was adjusted to maintain a dimensionless wall distance (y^+) close to unity, thereby accurately capturing boundary-layer behavior without relying on wall functions. Mesh independence was verified through successive grid refinement, ensuring that variations in lift and drag coefficients remained within acceptable numerical tolerance.

Turbulent effects were modeled using the k–ω SST turbulence model, which provides balanced performance for both near-wall treatment and free-stream flow prediction, particularly in cases involving adverse pressure gradients and potential flow separation [10-13]. Boundary conditions were defined as a pressure far-field inlet with specified Reynolds number values corresponding to the investigated cases, while a pressure outlet condition was applied at the downstream boundary. The airfoil surface was treated as a no-slip wall. Simulations were conducted under steady-state assumptions. Convergence was monitored through residual reduction criteria and stabilization of aerodynamic coefficients (CL and CD). Residuals were set to $10^{-7,}$ and parameters reached convergence. Since experimental validation of the hybrid Dolphin 0006 M4 configuration (Figure 1) is not included in the present study, the reliability of the obtained results is supported through consistency checks with published experimental and numerical data for comparable bio-inspired airfoils. Consequently, the CFD results are interpreted primarily in a comparative sense, focusing on relative performance differences between configurations rather than absolute aerodynamic prediction accuracy.

3. RESULTS

The analysis of selected bio-inspired airfoil optimization strategies indicates consistent improvements in aerodynamic performance across different design approaches. Both geometric modifications and biomimetic surface adaptations contribute to enhanced lift generation, reduced drag, and improved lift-to-drag ratio under specific flow conditions. A review of the literature shows that shark denticle-inspired surface structures can significantly increase aerodynamic efficiency, particularly at low angles of attack (below 4°) and Reynolds numbers on the order of 4×10^4. Reported improvements in lift-to-drag ratio reach up to 323% [15], while lift increases to

28.4% [10-13] and drag reductions of up to 18.09% [14] have also been documented for optimized configurations. The hybrid dolphin airfoils developed by the authors demonstrate improved aerodynamic characteristics compared to their baseline configurations. The results indicate consistent lift enhancement and drag reduction across the investigated operating range. Compared with selected bionic airfoil concepts in the literature, the hybrid dolphin configuration exhibits competitive aerodynamic performance. While certain literature-based optimization approaches show higher localized performance gains under specific conditions, the hybrid dolphin airfoils provide stable and balanced aerodynamic behavior across a broader range of operating regimes relevant to aerospace and defense applications. The Dolphin 0006 M4 airfoil was compared with ten bio-inspired airfoil concepts presented in Table 1. It should be emphasized once again that this comparison is intended only as a general comparison, since airfoils with different thickness ratios (e.g., 6%, 12%, 21%, etc.) cannot be directly compared.

Table 1. Comparative summary of optimized bionic airfoils and their aerodynamic improvements

Ref.	Bionic concept	Airfoil	Method	Re/flow velocity (m/s)	AoA (°)	ΔCL (%)	ΔCD (%)	ΔL/D (%)
[10–13]	**Hybrid dolphin airfoil (authors)**	**Dolphin 0006 M4**	**CFD**	**$6x10^6$**	**at 10**	**max. +28.4**	**min. −6.2**	**max. 41.9**
[14]	Bionic micro-grooves-shark inspired	NACA 0012	CFD	flow velocity 24-30	at 6	not reported	max. -18.09	max. +18.86%
[15]	Shark denticle-inspired surface	NACA 0012	CFD +Experiment	1×10^4 to 1×10^6	0-24	presented graphically	presented graphically	+323% at $Re=4\times10^4$, $\alpha<4^o$
[16]	Wavy leading edge airfoil	NACA63(4)-021	Experiment	2×10^5	0-90	max. +21	-3 (AoA: 30º-80º)	max. +23
[17]	Leading-edge tubercles	Wing model	Experiment	1.2 / 5.0 / 8.0×10^5	at 10	max. +18 (at $Re=5.0\times10^5$)	-2.72 (post-stall)	not reported
[18]	Leading-edge serrations	baseline wing with NACA0012	CFD	2×10^4	5-15	max. +15.92	drag increase	improved/ not reported
[19]	Bionic-inspired avian 2D airfoils	1. Long-eared owl airfoil 2. Sparrowhawk airfoil 3. Seagull airfoil	MIGA	1.02×10^5 to 1.45×10^6	(-5) to 20	lift increase	drag increase	increase (numerical values at different AoA, no specific %)
[20]	Tubercles on wing leading edge inspired by humpback whale flippers	NACA 643-221	CFD +Experiment	wind tunnels operating at speeds of up to 45	0-17	+4.8	-10.9	+17.6
[21]	Leading-edge tubercles	ONERA M6 based wing	CFD	high Re/ not specified	/	+16.7	/	+9
[22]	53 bionic airfoils with humpback whales' flipper tubercles	NACA 0009	Experiment	/	4-11	max. +28	/	/
[23]	Poro -Wavy Leading Edge	NACA 0012	CFD	4×10^5	0	Graphica-l	Graphica-l	/

The results obtained for the hybrid Dolphin airfoils are presented separately from the findings reported in the literature, which are subsequently considered for comparative analysis. Table 1 summarizes the main aerodynamic characteristics and reported performance improvements of the selected bio-inspired airfoil concepts, providing the basis for the comparative assessment of the hybrid Dolphin 0006 M4 against the selected literature-based configurations.

For clarification, the term baseline airfoil used in relation to the Dolphin 0006 M4 in this study refers to airfoil NACA 0006, although there are two baselines involved in this airfoil modification:

1. The original Dolphin airfoil developed by Iosif Taposu, as presented in Refs. [10-13]. The semi-elliptical ("natural") modification method, described in these studies, resulted in performance improvements relative to the original Dolphin configuration.

2. The NACA 0006 geometric counterpart is the primary baseline airfoil in terms of the aerodynamic improvements reported for the Dolphin 0006 M4 in Table 1 (first row). Although the original Dolphin 0006 developed according to Taposu's method (Refs. [10-13]) may also be regarded as the baseline configuration from a design perspective, the corresponding percentage improvements relative to that airfoil are not presented here because its aerodynamic performance is substantially lower than that of the NACA 0006. Consequently, the results in Table 1 quantify the performance gains of the Dolphin 0006 M4 with respect to the NACA 0006, while the modified airfoil itself originates from the original Dolphin 0006 through the semi-elliptic ("natural") modification described in Refs. [10-13].

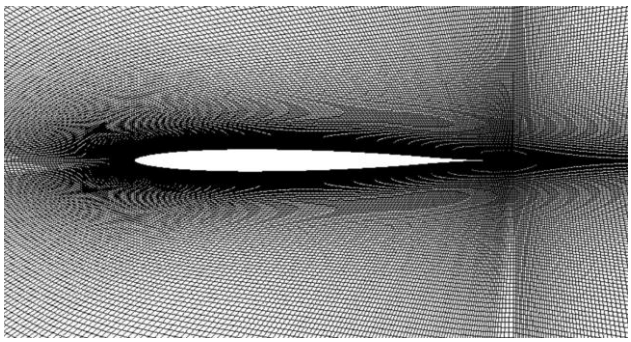


Figure 1. Computational mesh and geometry of the Hybrid Dolphin 006 M4 airfoil

Figure 1 presents the computational mesh and the geometry of the Hybrid Dolphin 0006 M4 airfoil with the corresponding structured C-mesh used for the CFD analysis performed by the authors of this paper. The mesh is refined in the vicinity of the airfoil surface to properly resolve boundary layer effects and ensure accurate prediction of aerodynamic quantities. The Dolphin 0006 M4 was designed using the semi-elliptical, or "natural" optimization method [10-13].

In addition to the quantitative values summarized in Table 1, the comparative analysis reveals distinct performance trends associated with different categories of bio-inspired optimization strategies. Surface-modification techniques, particularly those inspired by shark denticles and micro-groove structures, tend to exhibit pronounced improvements in aerodynamic efficiency at lower Reynolds numbers. These effects are primarily attributed to delayed boundary layer separation and modified near-wall momentum exchange mechanisms, which contribute to significant drag reduction under specific flow conditions.

In contrast, geometry-based modifications, such as leading-edge tubercles and wavy airfoil configurations, demonstrate a more balanced influence on both lift generation and flow stability. These configurations generally provide moderate but consistent improvements across a wider range of angles of attack, suggesting improved robustness rather than isolated peak performance gains.

The hybrid Dolphin 0006 M4 airfoil follows a similar trend to geometry-driven optimization approaches, where performance enhancement is distributed across the operating envelope rather than concentrated at specific flow conditions. The observed increase in lift coefficient and reduction in drag indicate favorable pressure distribution over the airfoil surface, which can be associated with smooth curvature transitions and optimized thickness distribution.

Unlike certain surface-texture-based configurations that exhibit extremely great localized improvements under narrowly defined conditions, the hybrid Dolphin configurations maintain stable aerodynamic behavior across the investigated operating range. This characteristic is particularly relevant for aerospace and defense applications, where operational variability and off-design conditions are critical considerations.

Furthermore, the comparison highlights a clear dependency of aerodynamic enhancement mechanisms on Reynolds number regime. At lower Reynolds numbers, viscous effects dominate, making surface texture modifications more effective. At higher Reynolds numbers, inertial effects become more pronounced, increasing the relative importance of global geometric optimization.

Overall, the results indicate that no single bio-inspired strategy universally outperforms others across all conditions. Instead, each approach demonstrates distinct advantages depending on the flow regime and design objective, with hybrid configurations offering a balanced compromise between performance improvement and operational stability.

4. DISCUSSION

The comparative analysis presented in Table 1 demonstrates that bio-inspired aerodynamic concepts can improve lift generation, reduce drag, and enhance aerodynamic efficiency through different physical mechanisms. However, direct comparison of reported improvements remains challenging because the reviewed studies employ different airfoil geometries, Reynolds number regimes, angles of attack, and evaluation

methodologies.

The Hybrid Dolphin 0006 M4 airfoil achieved a maximum lift increase of 28.4%, a drag reduction of 6.2%, and an L/D improvement of 41.9% at $Re = 6 \times 10^6$ and $\alpha = 10°$ [10–13]. Among the reviewed concepts, these results represent one of the highest simultaneously reported improvements in lift and aerodynamic efficiency under clearly defined operating conditions. However, a limitation of the present assessment is that the reported Dolphin 0006 M4 results are restricted to $Re = 6 \times 10^6$ and $\alpha = 10°$, which limits direct assessment of its performance across a broader range of operating conditions.

Several literature studies reported higher localized performance gains under specific conditions. Shark denticle-inspired surface modifications achieved L/D improvements of up to 323% at $Re \approx 4 \times 10^4$ and $\alpha < 4°$ [15], while shark-inspired micro-groove configurations reduced drag by up to 18.09% [14]. Wavy leading-edge airfoils increased lift by up to 21% [16], whereas leading-edge tubercle configurations reported lift improvements ranging from 4.8% to 28% depending on geometry and operating conditions [17, 20–22].

Nevertheless, many studies provide results only in graphical form or without explicit percentage quantification, which limits direct quantitative comparison [15, 17, 19, 21–23]. Furthermore, substantial differences in Reynolds number and AoA ranges prevent the establishment of a universal ranking among bio-inspired concepts.

The principal contribution of the present study is therefore not the claim of the greatest aerodynamic improvement, but the provision of a systematic and previously validated framework for evaluating a complete Hybrid Dolphin airfoil family. Unlike many studies focused on isolated geometric or surface modifications, the proposed approach enables consistent comparison within a unified design methodology. Such a framework is particularly relevant for aerospace and defense applications, where reproducibility, robustness, and reliable performance assessment are often more important than isolated peak improvements achieved under narrowly defined operating conditions. An important observation from the comparative analysis is the trade-off between peak aerodynamic performance and operational robustness. While some bio-inspired configurations achieve very great localized improvements in lift-to-drag ratio, these gains are often limited to narrow Reynolds number ranges and small angles of attack, which reduces their applicability in variable operating conditions. In contrast, the hybrid Dolphin 0006 M4 airfoil shows more stable aerodynamic behavior across a wider range of conditions. This suggests that its geometry supports smoother pressure distribution and reduced sensitivity to flow variations.

From an application perspective, this distinction is relevant for aerospace and defense systems, where consistency under off-design conditions is often more critical than isolated performance peaks. The results therefore indicate that hybrid bio-inspired designs offer a balanced compromise between efficiency and robustness.

5. CONCLUSION

This study presented a comparative evaluation of the hybrid Dolphin 0006 M4 against selected bio-inspired aerodynamic concepts reported in the literature. The analysis was based on aerodynamic performance indicators including lift coefficient, drag coefficient, and lift-to-drag ratio, obtained from numerical simulations, and experiments obtained from literature-based data.

The results show that the Dolphin 0006 M4 airfoil achieves a maximum lift increase of 28.4%, a drag reduction of 6.2%, and a lift-to-drag ratio improvement of 41.9% relative to its NACA 0006 baseline configuration. While certain bio-inspired surface modification strategies demonstrate higher localized performance gains under specific flow conditions, these improvements are often restricted to narrow operating regimes.

In contrast, the hybrid Dolphin 0006 M4 configuration exhibits stable aerodynamic behavior across a broader range of operating conditions, indicating improved robustness and practical applicability. This balance between performance enhancement and operational stability is particularly relevant for aerospace and defense applications, where reliability under variable flow conditions is essential.

The study confirms that hybrid bio-inspired design approaches, supported by computational optimization techniques, provide a viable framework for achieving consistent aerodynamic improvements. Future work may focus on further refinement of hybrid geometries and multi-objective optimization strategies to enhance performance across extended flight regimes.

Acknowledgments

This research is supported by the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia under Contract No. 451-03-33/2026-03/200213, dated February 5, 2026.

References

[1] YAN, H., SU, X., ZHANG, H., LIU, X.: *Design approach and hydrodynamic characteristics of a novel bionic airfoil*, Ocean Engineering, 216 (2020) 108076. https://doi.org/10.1016/j.oceaneng.2020.108076

[2] CHEN, L., YANG, X., YAN, H., LIU, X.: *Optimal design and application of bionic hydrofoil with multiple angles of attack based on genetic algorithm*, Ocean Engineering, 239(18) (2025). https://doi.org/10.1177/09544062251347112

[3] LOYD, C. J., PEAKALL, J., BURNS, A. D., KEEVIL, G. M., DORRELL, R. M., WIGNALL, P.B., FLETCHER, T.M.: *Hydrodynamic efficiency in sharks: the combined role of riblets and denticles*, Bioinspiration & Biomimetics, 16(4) 046008 (2021). https://doi.org/10.1088/1748-3190/abf3b1

[4] BLIAMIS, C., VLAHOSTERGIOS, Z., MISIRLIS,

D., YAKINTHOS, K.: *Modeling surface riblets skin friction reduction effect with the use of computational fluid dynamics*, Chemical Engineering Transactions, 81 (2020) 595–600. https://doi.org/10.3303/CET2081100

[5] GHOSH, A., GUPTA, P., JAYANT, SINGH, R. K.: *Metaheuristic optimization framework for drag reduction using bioinspired surface riblets*, arXiv preprint, Cornell University (2021). https://doi.org/10.48550/arXiv.2109.09650

[6] MOHAMMADIKARACHI, A., YOUSIF, M. Z., NUGROHO, B., LIM, H. C.: *An exhaustive review of studies on bio-inspired convergent-divergent riblets*, arXiv preprint, Cornell University (2023). https://doi.org/10.48550/arXiv.2309.09138

[7] GRAYBILL, M.T., XU, N.W.: *Experimental studies of bioinspired shark denticles for drag reduction*, Integrative and Comparative Biology, 64(3) (2024) 742-52. https://doi.org/10.1093/icb/icae086

[8] GOLDBERG, D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Addison-Wesley Longman Publishing Co., Inc. Boston, 1989.

[9] WU, A., MA, T., WANG, Sh., DING, Ch.: *Optimization of Multi-Objective Problems for Sailfish-Shaped Airfoils Based on the Multi-Island Genetic Algorithm*, Machines, 13(8) (2025) 637. https://doi.org/10.3390/machines13080637

[10] DANČUO, Z. Z., KOSTIĆ, I. A., KOSTIĆ, O. P., BENGIN, A. Č., VOROTOVIĆ, G. S.: *Initial Development of the Hybrid Semielliptical-Dolphin Airfoil*, Thermal Science, 26(3) (2022) 2199-210. https://doi.org/10.2298/TSCI210515234D

[11] DANČUO, Z. Z.: *Development of a Family of Hybrid Dolphin Airfoils*, Doctoral Thesis, University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia, 2023. Available at: https://hdl.handle.net/21.15107/rcub_machinery_7672

[12] DANČUO, Z. Z., KOSTIĆ, I. A., KOSTIĆ, O. P., BENGIN, A. Č., VOROTOVIĆ, G. S.: *Influence of Thickness Ratio on the Aerodynamic Characteristics of a Family of Hybrid Semielliptical Dolphin Airfoils*, in: Novel Techniques in Maintenance, Repair, and Overhaul. Sustainable Aviation, Springer Nature (2024) 1-14. https://doi.org/10.1007/978-3-031-42041-2_1

[13] DANČUO, Z. Z., KOSTIĆ, I. A., KOSTIĆ, O. P., BENGIN, A. Č., VOROTOVIĆ, G. S.: *Experimental Study on the Aerodynamics of a Hybrid Dolphin Airfoil*, Tehnički vjesnik – Technical Gazette, 33(2) (2026) 675-83. https://doi.org/10.17559/TV-20250607002729

[14] WU, Z., YANG, Y., LIU, M., ZHANG, K.: *Numerical Study on Flow Characteristics of Airfoil With Bionic Micro-Grooves*, Research Square Preprint, (2020). https://doi.org/10.21203/RS.3.RS-88806/V1

[15] DOMEL, A. G., SAADAT, M., WEAVER, J. C., HAJ-HARIRI, H., BERTOLDI, K., LAUDER, G. V.: *Shark Skin-Inspired Designs That Improve Aerodynamic Performance*, Journal of the Royal Society Interface, 15(139) (2018). http://dx.doi.org/10.1098/rsif.2017.0828

[16] WANG, G. F., ZHANG, M. M., XU, J. Z.: *Experimental Study on Bionic Wavy Edge Aerodynamic Characteristics of Airfoil*, Journal of Engineering Thermophysics, 34(10) (2013) 1842–1846.

[17] ÇOLAK, A., SEYHAN, M., SARIOĞLU, M.: *Leading-Edge Tubercle Modifications to the Biomimetic Wings*, Physics of Fluids, 35(1) (2023). https://doi.org/10.1063/5.0131803

[18] LIU, K., SONG, B., XUE, D., YANG, W., CHEN, A., WANG, Z.: *Numerical Study of the Aerodynamic Effects of Bio-Inspired Leading-Edge Serrations on a Heaving Wing at a Low Reynolds Number*, Aerospace Science and Technology, 124 (2022) 107529. https://doi.org/10.1016/j.ast.2022.107529

[19] ZHANG, Z., WU, B., WU, L., LIU, W., LIU, L., LI, N., CUI, L.: *Optimization of the Bionic Wing Shape of Tidal Turbines Using Multi-Island Genetic Algorithm*, Machines, 11(1) (2023) 43. https://doi.org/10.3390/machines11010043

[20] MOHAMED, H., EL-SHAMY, A., AL-ANSARI, D. E., ABDELAZEM, Y. N., HABIB, M. R., HASSANEIN, M. A., EL-BAHLOUL, S. A., MOUSTAFA, W. S.: *Investigating the Impact of Tubercle Modifications on the Performance of the NACA 643-221 Airfoil*, Journal of Physics: Conference Series, 3070(1) (2025) 012011. https://doi.org/10.1088/1742-6596/3070/1/012011

[21] GUO, X., XU, J., HUANG, Y., CAO, D.: *Aerodynamic Characterization of Leading-Edge Tubercles Bionic Airfoils with Different Configurations at High Reynolds Number*, Physics of Fluids, 36(12) (2024) 125140. https://doi.org/10.1063/5.0243806

[22] FISH, F. E., WEBER, P. W., MURRAY, M. M., HOWLE, L. E.: *The Tubercles on Humpback Whales' Flippers: Application of Bio-Inspired Technology*, Integrative and Comparative Biology, 51(1) (2011) 203–213. https://doi.org/10.1093/icb/icr016

[23] CHEN, W., LEI, H., XING, Y., WANG, L., ZHOU, T., QIAO, W.: *On the Airfoil Leading-Edge Noise Reduction Using Poro-Wavy Leading Edges*, Physics of Fluids, 36(3) (2024). https://doi.org/10.1063/5.0198034

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF UNDERWING PAYLOAD ON THE FLIGHT PERFORMANCE OF A TACTICAL UAV

BRANISLAV OSTOJIĆ
Military Technical Institute, Belgrade, branislavostojic1994@gmail.com

MIODRAG MILENKOVIĆ-BABIĆ
Military Technical Institute, Belgrade, miodragmbm@yahoo.co.uk

ALJOŠA LJUBOJA
Military Technical Institute, Belgrade, aljosabiljana@gmail.com

OGNJEN PEKOVIĆ
Faculty of Mechanical Engineering, Belgrade, opekovic@mas.bg.ac.rs

Abstract: *This paper presents a performance analysis of a tactical unmanned aerial vehicle (UAV) with and without integrated rocket payload. The study is based on the application of standard flight mechanics methods combined with available experimental flight data for validation purposes. The aerodynamic characteristics of the UAV are evaluated for both configurations, taking into account the increase in mass and additional parasitic and interference drag caused by externally mounted rockets. These changes are incorporated into the equations of motion in order to estimate key performance parameters. The analysis focuses on the determination of maximum flight speed, rate of climb, endurance, and range under different flight conditions. Two calculation methodologies are applied and their results are compared with experimental data obtained during flight testing. The results show a noticeable degradation of flight performance due to rocket integration, with deviations between theoretical predictions and experimental data remaining within acceptable engineering limits. The presented approach provides a reliable basis for performance estimation of UAVs with external payload configurations.*

Keywords: *UAV, rocket payload, flight performance, flight testing.*

1. INTRODUCTION

The development of tactical unmanned aerial vehicles began during the second half of the 20th century, primarily for reconnaissance and intelligence-gathering purposes without risking human crews. During military conflicts worldwide, UAV systems were significantly improved, evolving from simple reconnaissance platforms into complex systems with greater autonomy, reliability, and operational range. At the beginning of the 21st century, intensive technological development in the fields of sensors, communication systems, and precision-guided weaponry enabled the transformation of UAV systems into multifunctional platforms. Modern tactical unmanned aerial vehicles, developed, among others, at the Military Technical Institute, represent the integration of reconnaissance and combat capabilities, offering a high degree of flexibility and relatively low operational costs. Such development further confirms their increasingly important role in modern military operations.

Medium-range unmanned aerial vehicles, such as the system considered in this paper, are designed as platforms capable of performing a wide range of tasks, including surveillance, intelligence gathering, artillery fire correction, and precision strike missions. Their ability to remain in the mission area for extended periods and transmit data in real time significantly enhances situational awareness and command efficiency. The integration of external stores further expands the operational capabilities of these systems; however, it also affects the aircraft's fundamental aerodynamic and mass characteristics. As demonstrated in previous studies, any modification of the configuration requires careful analysis of its impact on the center of gravity position, drag, and overall performance, while achieving a balance between operational efficiency and flight performance [1-3].

The integration of rocket armament enables the execution of direct fire support and reduces reaction time, thereby combining reconnaissance and combat functions within a single platform. However, the presence of external stores leads to an increase in mass and parasite drag, as well as the occurrence of additional interference drag, which directly results in reduced maximum speed, climb rate, endurance, and range. In addition, limitations on the maximum takeoff weight often require a reduction in fuel quantity, while changes in mass distribution may affect the aircraft's stability and controllability. Although aerodynamically unfavorable, such a configuration represents an operationally justified compromise in missions where combat capability takes priority over maximum flight performance.

In this context, there is a need for a reliable assessment of

Corresponding author: B. Ostojić;
E-mail: branislavostojic1994@gmail.com
DOI: 10.5937/OTEH26002O

the influence of external stores on the performance of unmanned aerial vehicles. Although standard analytical methods based on the literature exist [4-6], their application to configurations with external payloads requires additional verification.

The aim of this paper is to provide a quantitative analysis of the impact of rocket armament integration on the performance of an unmanned aerial vehicle through a combination of theoretical calculations and experimental data. The analysis includes the determination of drag, required power, maximum speed, climb rate, endurance, and range, together with a comparison to the results obtained from flight tests conducted in configurations with and without rockets. Special emphasis is placed on evaluating the deviations between theoretical models and actual performance, as well as identifying the dominant factors contributing to performance degradation caused by the integration of external payloads. The obtained results are used to assess the accuracy and limitations of existing UAV performance calculation methods for configurations with external stores (rockets) through comparison with flight test results.

2. FLIGHT PERFORMANCE

The focus of this chapter is the analysis of the UAV flight performance in configurations with and without externally mounted rocket armament, with the aim of assessing the influence of external payloads on the main flight characteristics. The performance analysis was carried out using standard analytical methods for determining drag, required power, maximum speed, climb rate, range, and endurance. The obtained results will be compared with experimental data from flight tests in order to evaluate the accuracy of the applied methods and identify the dominant factors influencing UAV flight performance.

Table 1 presents some of the basic characteristics of the considered unmanned aerial vehicle shown in Figure 1.

Table 1. Key Parameters of the UAV Configuration

Internal combustion engine power [7]	38.8 kW
Propeller diameter	0.86 m
Maximal mass of UAV	265 kg
Maximal mass of usable fuel	70 kg
Wing area	4.33 m^2
Mean aerodynamic chord	0.736 m

Figure 1 Tactical UAV with rockets

The propulsion system must be properly sized to ensure the overcoming of aerodynamic drag and supplying power to all onboard systems. In this regard, it is necessary to take into account that different engines have specific optimal operating RPM ranges. The efficiency of the propeller, as an integral part of the propulsion system, depends on propeller advance ratio, which is determined using the following expression:

$$J = \frac{V}{ND} \tag{1}$$

In Equation (1), the parameter J represents the propeller advance ratio, V the flight speed, N the propeller rotational speed, and D the propeller diameter. Maximum propeller efficiency is achieved when the propeller advance ratio is within the range of 0.5 to 0.6, as illustrated in Figure 2.

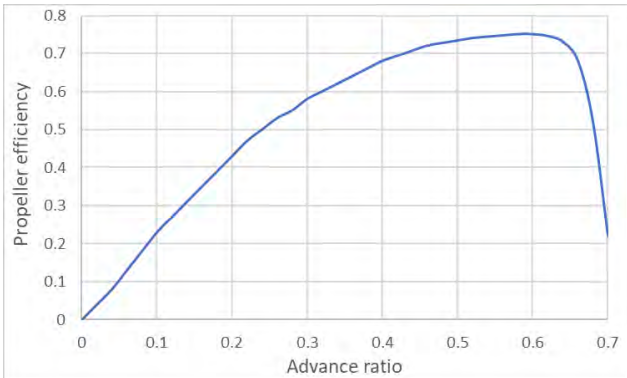


Figure 2. Propeller efficiency

For the flight speed to be optimal, it is necessary to satisfy several interrelated conditions. Depending on the mission type, the UAV must achieve an appropriate flight speed at the designated altitude in order to ensure efficient terrain surveillance, tracking of moving and stationary targets, as well as successful rocket launch and target engagement. During different flight regimes, the UAV changes its configuration, including camera deployment and retrieval, as well as variations in the amount of carried armament, which lead to changes in mass and aerodynamic characteristics, and consequently to variations in the required flight power, as shown in Figure 3 and Figure 4.

Figure 3 shows the relationship between the lift coefficient C_L and the drag coefficient C_D for several different configurations (with and without rockets) [8].

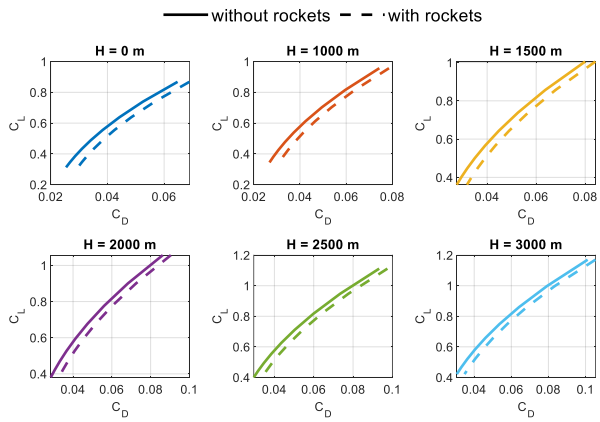


Figure 3. UAV drag polar with and without rockets

Based on the obtained drag coefficient values, the aerodynamic drag force and required flight power can be calculated using the following expressions:

$$D = C_D q S \tag{2}$$

$$P_{req} = DV \tag{3}$$

Figure 4 shows the influence of rocket integration on the required flight power, as well as the variation in available power. Figure 5 presents the maximum flight speeds as a function of altitude for configurations with and without rockets.

It should be emphasized that engine power depends on the rotational speed, throttle setting, flight regime, and atmospheric conditions. With increasing altitude, air pressure, density, and temperature decrease, resulting in reduced engine efficiency. One of the methods used to express the available engine power as a function of altitude is given by:

$$P_{avail} = P_0 \eta_{engine} \eta_{prop} \left(\frac{\rho}{\rho_0}\right)^{1.2} \tag{4}$$

where ρ is the air density at the flight altitude, while ρ_0 is the air density at sea level.

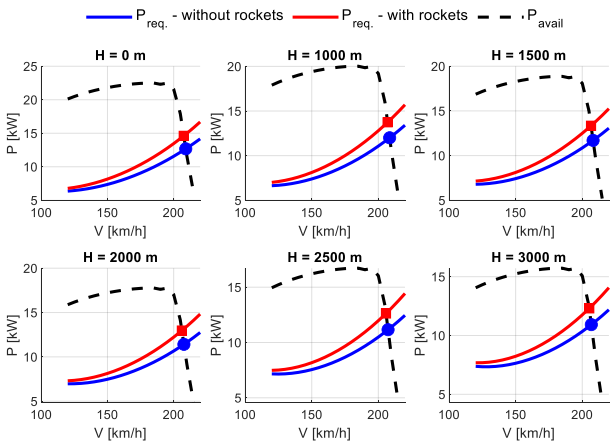


Figure 4. Required and available flight power

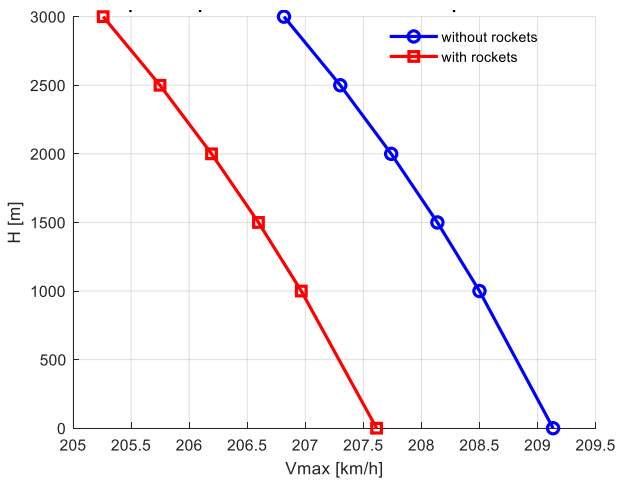


Figure 5. Maximum flight Speed as a Function of Altitude

The vertical rate of climb depends on the difference between the available and required flight power, as well as on the current aircraft weight:

$$w = \frac{P_{avail} - P_{req}}{mg} \tag{5}$$

Figure 6 illustrates the influence of rocket integration on the rate of climb at different flight speeds and altitudes.

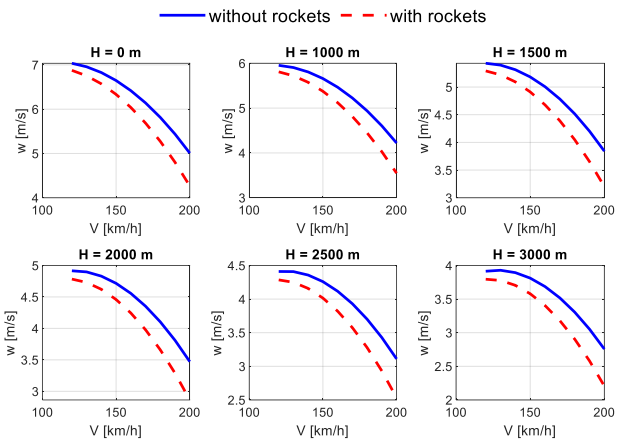


Figure 6. Rate of climb as a function of airspeed and altitude

As shown in Figure 7, with increasing flight altitude, the rate of climb gradually decreases, and for certain flight regimes and airspeeds, the climb rate approaches zero. Such behavior is a consequence of the reduction in available power and thrust with increasing altitude, as well as the increase in the power required to maintain flight under specific aircraft configurations. A similar decrease in climb rate with increasing altitude was also reported in [9] for a propeller-driven UAV operating at altitudes up to 7000 m. Depending on the mission requirements, aerodynamic configuration, and required aircraft performance, the service ceiling may be defined at certain minimum climb rate values, while the theoretical ceiling corresponds to the condition in which the rate of climb reaches 0 m/s.

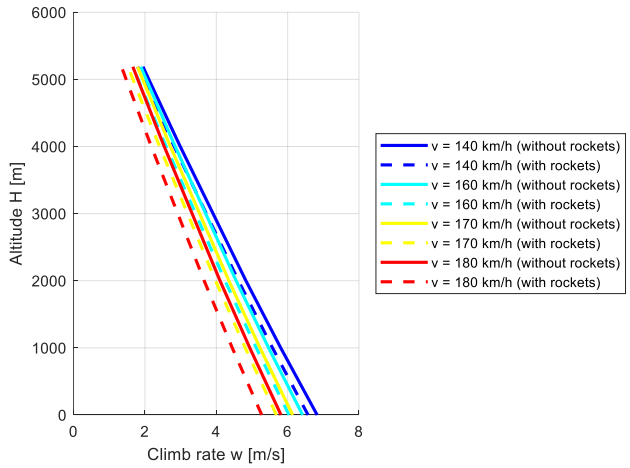


Figure 7. Rate of climb as a function of altitude

When considering flight endurance, the primary objective in the design and analysis of an unmanned aerial vehicle is to ensure the longest possible time aloft without the need for landing and refueling, which requires the optimization of multiple interdependent aerodynamic, propulsion, and mass-related parameters. In this context, one of the key requirements is achieving a high lift-to-drag ratio (L/D),

which directly reduces the power required for sustained level flight, while minimum drag is achieved through optimization of the aircraft configuration and external stores. At the same time, low specific fuel consumption of the propulsion system is of crucial importance, as it enables more efficient utilization of the available energy through the selection of an engine operating within its optimal regime and providing a favorable balance between power output and fuel consumption. The propeller efficiency, as a key element of the propulsion system, must also be maximized in order to convert the largest possible portion of the engine's mechanical power into useful thrust. In addition, the total amount of fuel that the UAV can carry represents a significant factor, although it is constrained by structural and operational requirements such as maximum takeoff weight and load distribution. Therefore, optimal endurance is achieved through a careful balance between aerodynamic efficiency, propulsion system performance, and available energy resources, while taking into account all limitations of the actual operational environment.

Depending on the flight regime, range and endurance can be classified into the following cruise conditions:

- Constant Airspeed/Constant Altitude Cruise.
- Constant Attitude/Constant Altitude Cruise.
- Constant Airspeed/Constant Attitude Cruise.
- Constant Weight Cruise.

This paper presents range and endurance values for steady flight at constant speed and constant altitude, without considering takeoff, climb to cruise altitude, descent, and landing phases. Table 2 presents the corresponding range and endurance expressions, which yield identical results.

Table 2. Analytical expressions for range and endurance

Range [10]	$R = 3600\frac{\eta}{q}\sqrt{\frac{\pi\lambda_e}{C_{Dmin}}}\arctan\frac{m_g}{\frac{\delta V^2}{16}S\sqrt{C_{Dmin}\pi\lambda_e} + \frac{m_o(m_o - m_g)}{\frac{\delta V^2}{16}S\sqrt{C_{Dmin}\pi\lambda_e}}}$
Endurance [10]	$E = \frac{R}{V}$
Range [11]	$R = \frac{V}{c_t\sqrt{kC_{Dmin}}}\left[tan^{-1}\left(\frac{2\sqrt{k}}{\rho V^2 S\sqrt{C_{Dmin}}}W_{ini}\right) - tan^{-1}\left(\frac{2\sqrt{k}}{\rho V^2 S\sqrt{C_{Dmin}}}W_{fin}\right)\right]$
Endurance [11]	$E = \frac{1}{c_t\sqrt{kC_{Dmin}}}\left[tan^{-1}\left(\frac{2\sqrt{k}}{\rho V^2 S\sqrt{C_{Dmin}}}W_{ini}\right) - tan^{-1}\left(\frac{2\sqrt{k}}{\rho V^2 S\sqrt{C_{Dmin}}}W_{fin}\right)\right]$

Figure 8 and Figure 9 present the range and endurance of the analyzed UAV for different flight regimes in configurations with and without rockets. The results show that the UAV configuration with rockets has lower range and endurance compared with the configuration without rockets. This reduction is mainly attributed to the decrease in available fuel capacity caused by rocket instalation, which limits both the total flight time and the achievable range.

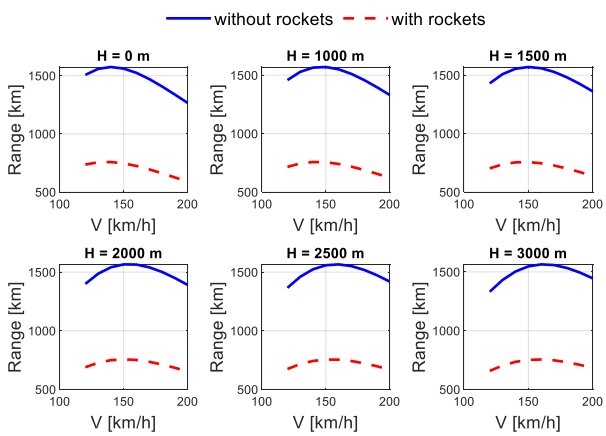


Figure 8. Range as a function of airspeed

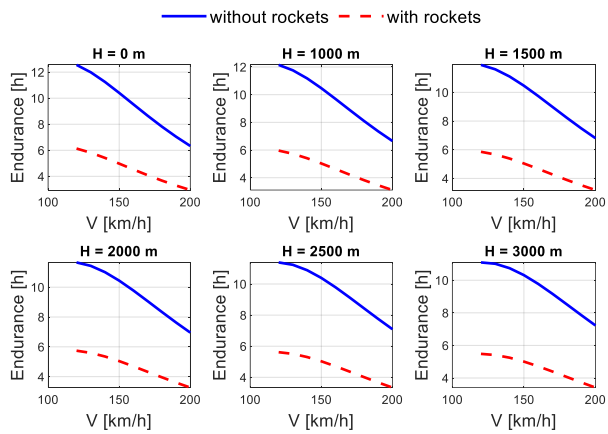


Figure 9. Endurance as a function of airspeed

2.1. Flight test results

The maximum flight speed was evaluated following stabilization at the cruise altitude of 2000 m, where the UAV was operated at a throttle setting of 90% for a duration of three minutes. The UAV was then climbed to an altitude of 3000 m, and the same test procedure was repeated. In order to minimize the influence of wind on the

measurement results, the tests were conducted in both flight directions at each altitude. Table 3 provides a comparison between the calculated performance values and the results obtained during flight testing. It should be noted that the objective of the flight tests was not to determine the maximum attainable flight speed, but rather to verify compliance with the required operational speeds specified in the technical requirements, namely 180 km/h for the configuration without rockets and 160 km/h for the configuration with rockets.

Table 3. Comparison of calculated and measured maximum flight speeds

Without rockets		
H (m)	$V_{calculated}$ (km/h)	V_{test} (km/h)
2000	207.74	190.8
3000	206.82	192.6
With rockets		
H (m)	$V_{calculated}$ (km/h)	V_{test} (km/h)
2000	206.19	186.01
3000	205.26	187.09

Table 4 presents a comparison between the calculated and experimentally measured rates of climb. For the configuration without rockets, the deviations between the calculated and measured values were below 10%, with a maximum difference of 8.1%, indicating good agreement between the analytical model and flight test results. Larger deviations were observed for the configuration with rockets. These discrepancies are attributed to differences in the conditions under which the flight tests were conducted, since the flight tests in the configuration with rockets and the flight tests in the configuration without rockets were performed on different days. As a result, variations in atmospheric conditions, including turbulence encountered during cloudy weather, affected the measured rates of climb.

Table 4. Comparison of maximum rates of climb

Without rockets		
H (m)	$w_{calculated}$ (m/s)	w_{test} (m/s)
1000	5.95	6.33
2400	4.92	4.52
With rockets		
H (m)	$w_{calculated}$ (m/s)	w_{test} (m/s)
1000	5.77	6.62
2200	4.82	6.78

The service ceiling was evaluated by climbing the UAV to an altitude of 5000 m above sea level in the configuration without rockets and to 4000 m in the configuration with rockets, while maintaining a stable flight condition for a duration of three minutes in both cases. The calculated dependence of rate of climb on altitude is presented in Figure 7, while Table 5 presents the corresponding flight test results. It should be noted that the objective of the flight tests was not to determine the maximum attainable altitude, but rather to verify compliance with the required operational ceiling specified in the technical requirements, namely 5000 m for the configuration without rockets and 4000 m for the configuration with rockets. Consequently, the measured values differ from the calculated results, particularly in the configuration with rockets.

Table 5. Service ceiling comparison

H (m)	$w_{cal.}$ (m/s)	w_{test} (m/s)	Rockets
5188	1.95	2	without
4029	2.3	1	with

Endurance was one of the critical requirements that the aircraft had to satisfy during flight testing. Since the UAV successfully met the required flight endurance criteria, maximum range testing was not performed. In order to determine the range and endurance, it is first necessary to evaluate the fuel consumption during ground engine operation, climb, descent, landing, and taxiing. When the required navigation reserve is additionally considered, the remaining available fuel quantity for steady level flight at the prescribed cruise altitude can be determined.

In this study, the endurance values were compared exclusively for the flight segment extending from reaching the target flight altitude to the initiation of descent, with the flight conducted at an average throttle setting of 30%.

During the flight tests conducted in the configuration without rockets, the unmanned aerial vehicle remained airborne for 9.42 hours at an average cruise altitude of 2500 m and an average flight speed of 150 km/h, consuming 46 kg of fuel, as shown in Figure 10. The calculated endurance for the corresponding flight regime was 10.14 hours.

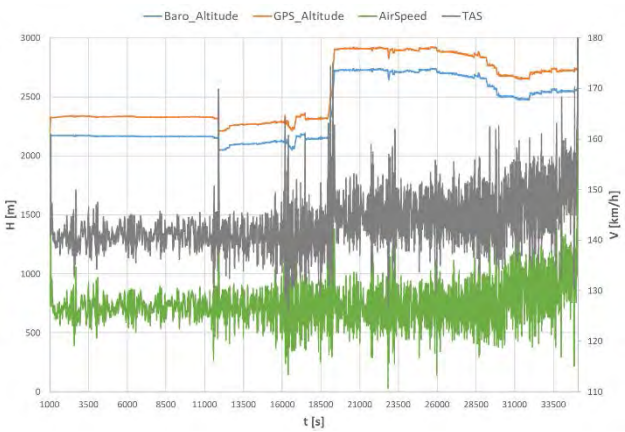


Figure 10. UAV Endurance in the configuration without rockets [12]

During the flight tests conducted in the configuration with rockets, the unmanned aerial vehicle remained airborne for 5.39 hours at an average cruise altitude of 2000 m and an average flight speed of 140 km/h, consuming 26 kg of fuel, as shown in **Figure 11**. The calculated endurance for the corresponding flight regime was 5.35 hours.

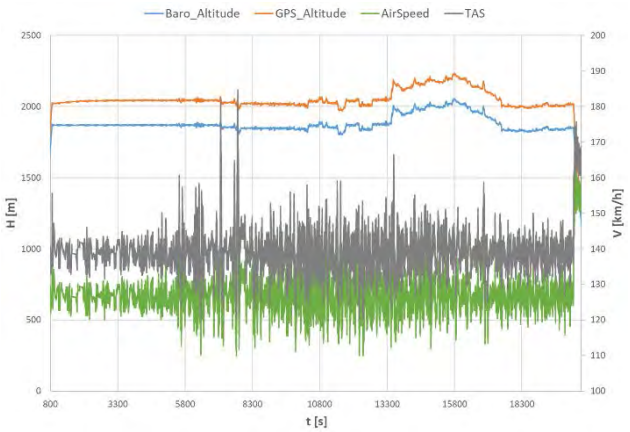


Figure 11. UAV Endurance in the configuration with rockets [12]

3. CONCLUSION

The paper presented a comparative analysis of the flight performance of a tactical unmanned aerial vehicle in configurations with and without external rocket stores. The analysis included the determination of maximum flight speed, rate of climb, endurance, and range through the application of standard analytical flight performance methods, followed by comparison with the results obtained during flight testing.

The obtained results indicate that rocket integration has a relatively small influence on maximum flight speed and rate of climb. Although the presence of external stores leads to an increase in aerodynamic drag and aircraft mass, the corresponding degradation of these performance parameters remained comparatively low under the analyzed flight conditions. In contrast, a significantly greater impact was observed in terms of endurance and range. The primary reason for this reduction was not the increase in aerodynamic drag itself, but rather the decrease in available fuel quantity required to satisfy the maximum takeoff mass limitation in the armed configuration.

A comparison between the calculated and experimentally obtained results showed good agreement across all analyzed cases, confirming that conventional analytical flight performance methods provide sufficiently accurate predictions even for UAV configurations with external stores. This demonstrates that relatively simple analytical approaches can still serve as reliable tools during preliminary design and operational performance assessment.

It should also be noted that flight testing of UAV performance is inherently influenced by atmospheric conditions, which may vary significantly during the execution of the tests. Sudden wind gusts and local atmospheric disturbances can affect the measured flight parameters, potentially contributing to deviations between calculated and experimentally obtained results, even under carefully controlled test conditions.

Future research may include the investigation of advanced endurance enhancement methods for tactical UAVs, such as autonomous soaring, atmospheric energy harvesting, and energy-efficient trajectory optimization techniques presented in recent studies [13].

Finally, despite the reduction in endurance and range, the capability to integrate rocket armament represents a major operational advantage. Modern combat operations increasingly emphasize the importance of unmanned platforms capable of combining reconnaissance and strike functions within a single system, making armed UAV systems an increasingly important component of contemporary military operations.

Acknowledgment

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03-34/2026-03/200325).

References

[1] OSTIĆ,B., MILENKOVIĆ-BABIĆ, M., STEFANOVIĆ-GOBELJIĆ, V., OSTOJIĆ, B.: *Analysis of the mass and aerodynamical changes due to implementation of an external rescue system on a medium range UAV*, OTEH 2024, Belgrade, 2024.

[2] STEFANOVIĆ,V., MILENKOVIĆ-BABIĆ, M., ZDRAVKOVIĆ M.: *Analysis of the rescue system implementation in the medium range UAV*, OTEH 2024, Belgrade, 2024.

[3] ANTONIĆ, V., TRIFKOVIĆ, M., MOLOVIĆ, V., MILENKOVIĆ-BABIĆ, M.: Small Fixed-Wing UAV Precision Aerial Drop Capability Development, Scientific Technical Review, 74(2) (2024) 65-70.

[4] MILENKOVIĆ-BABIĆ, M., DOVATOV, B., OSTOJIĆ, B., ANTONIĆ, V., TRIFKOVIĆ, M., IVKOVIĆ, D., Tactical UAV Flight Performance Estimation and Validation, Defence Science Journal, Vol. 75, No. 1, January 2025, pp. 19-26,

[5] ROSKAM, J., JAN, C., Airplane Aerodynamics and Performance, 1997.

[6] RAYMER, D., Aircraft Design: A Conceptual Approach, Sixth edition, AIAA education series, 2018.

[7] ZANZOTTERA TECHNOLOGIES S.r.l., 630HS Engines, https://www.zanzotteraengines.com/engines/630hs-engine/

[8] Proračun aerodinamičkih karakteristika taktičke bespilotne letelice srednjeg doleta -PEGAZ, primenom CFD-a, VTI, Beograd, 2020.

[9] LIU, R.-L., ZHANG, Z.-J., JIAO, Y.-F., YANG, C.-H., and ZHANG, W.-J. Study on flight performance of propeller-driven UAV, International Journal of Aerospace Engineering, Vol. 2019, Article ID 6282451, 2019.

[10] RENDULIĆ, Z., Mehanika leta, 1987.

[11] GUDMUNDSSON, S. General Aviation Aircraft design: Applied Methods and Procedures, first edition 2014.

[12] B3–0609–И–06, Izveštaj o zemaljskim i letnim ispitivanjima letelice, VTI, Beograd, 2022.

[13] LAWRANCE, N. R. J., ACEVEDO, J. J., CHUNG, J. J., NGUYEN, J. L., WILSON, D., and SUKKARIEH, S., Long endurance autonomous flight for unmanned aerial vehicles, AerospaceLab, Issue 8, December 2014, AL08-05.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SYSTEMATIC MODELING AND EXPERIMENTAL EVALUATION OF BLDC MOTOR-PROPELLER SYSTEMS IN MULTICOPTERS

ALJOŠA LJUBOJA
Military Technical Institute, Belgrade, aljosabiljana@gmail.com

MILENKO TRIFKOVIĆ
Military Technical Institute, Belgrade, milenkot@gmail.com

BRANISLAV OSTOJIĆ
Military Technical Institute, Belgrade, branislavostojic1994@gmail.com

TONI IVANOV
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, tivanov@mas.bg.ac.rs

Abstract: *This paper presents a systematic methodology for assessing the performance of a multicopter propulsion system, focusing on the propeller, brushless direct current (BLDC) motor, electronic speed controller (ESC), and battery system. Addressing the challenge that current literature and datasheets on UAV performances especially concerning electric motors and propellers are usually scattered and poor. This research provides specific motor characteristics that are often difficult to find. By utilizing manufacturer technical specifications, mathematical models are used to evaluate critical metrics, including hover endurance, motor–ESC efficiency across the throttle range, payload capacity and tilt limits, and the optimization of speed and range in forward flight. Central to this research is the comprehensive experimental validation conducted using a motor thrust stand. Detailed experimental measurements of static thrust, torque, propeller RPM, and electricity consumption were performed to validate the numerical models. The results demonstrate a correlation between theoretical predictions and experimental data, confirming that the proposed methodology allows for reliable estimation of flight performance. This integrated approach provides a robust framework for engineers to effectively optimize component selection and accurately predict multicopter behavior.*

Keywords: *Multicopter, Propulsion System, BLDC Motor, Experimental validation.*

1. INTRODUCTION

The quadcopter is the most common multicopter unmanned-aerial-vehicle configuration. Vertical take-off and landing, hovering, low mechanical complexity and precise positioning support reconnaissance, surveillance, relay, logistics and specialized payload missions. In military applications, low observability, short preparation time, and the ability to operate outside the operator's direct line of sight are particularly important [1-3]. At the same time, low available battery energy and limited payload nevertheless remain primary constraints.

Performance depends on matching the propeller, brushless direct-current (BLDC) motor, electronic speed controller (ESC) and battery. Manufacturers do not publish a uniform data set: motor KV, current and resistance may be available, whereas a propeller is often described only by diameter, pitch and blade count. Static thrust and torque coefficients are consequently uncertain, especially at low Reynolds number, while forward flight additionally changes loading through advance ratio and disk incidence [4]. Low-order component models are useful in preliminary design because they connect a small parameter set to hover, endurance and mission estimates [5-7]. For forward flight and maneuvering conditions, where rotors operate under non-uniform and oblique inflow, propeller maps or suitable aerodynamic models are required [8-11]. A calculation based on a zero-advance ratio static test must therefore be reported as a conditional engineering screen rather than as measured flight performance.

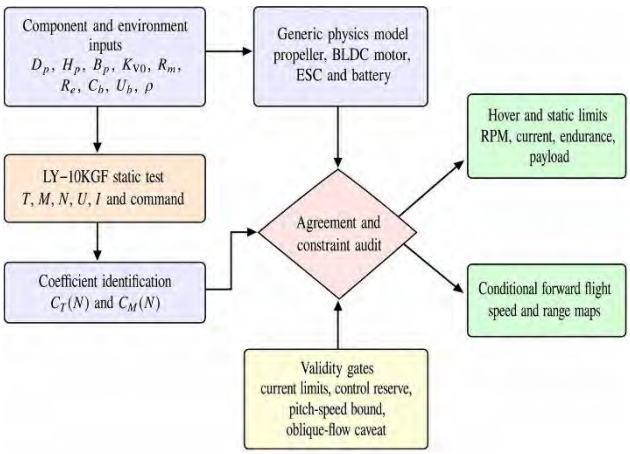


Figure 1. Workflow from confirmed hardware and LY-10KGF data to calibrated coefficients, constraint checks and conditional vehicle-level analysis

The objective is to determine operating points relevant to

Corresponding author: A. Ljuboja;
E-mail: aljosabiljana@gmail.com

DOI: 10.5937/OTEH26003L

hover and high loading, identify ESC and motor constraints, estimate payload, tilt, speed and range, and quantify agreement with a stationary thrust stand experiment. Four experimental measurements of a GEPRC EM3115 900 KV motor, 10x5x3-in propeller and HAKRC 65 A ESC were performed using an LY-10KGF thrust stand. After removing the 90% and 100% regimes, which will be explained later, 32 points remain; 20 points from 40% to 80% are used for coefficient identification. Figure 1 summarizes the traceable workflow.

2. SYSTEM AND METHODS

2.1. Tested configuration

The studied vehicle is m = 5.4 kg quadcopter. Each branch uses a GEPRC EM3115 900 KV motor [12] and a three-blade 10x5x3-in propeller shown in Figure 2. The used electronic speed controller is a HAKRC unit rated at 65 A per branch [13]. The shared source is a 6S, 10 Ah, 30 C LiPo battery. Air density is measured in each experiment and lies between 1.174÷1.175 kg/m³ at 23.2÷23.6 °C. Table 1 lists the model inputs.

Figure 2. Used 10x5x3-in propeller and GEPRC EM3115 900 KV motor

Table 1. Confirmed configuration and test environment

Item	Value
Vehicle	5.4 kg; four rotors
Motor	GEPRC EM3115, 900 KV
Propeller	10x5x3 in; three blades
ESC	HAKRC, 65 A per branch
Battery	6S LiPo; 10 Ah; 30 C;
Atmosphere	23.2÷23.6 °C; ρ=1.174÷1.175 kg m^{-3}

2.2. Analytical and identified static model

At zero advance ratio, thrust T and reaction torque M are expressed by Equations (1a) and (1b), where N is optical speed in RPM, D_p is propeller diameter, ρ is air density, C_T thrust coefficient and C_M is torque coefficient.

$$T = C_T \rho \left(\frac{N}{60}\right)^2 D_p^4 \tag{1a}$$

$$M = C_M \rho \left(\frac{N}{60}\right)^2 D_p^5 \tag{1b}$$

The geometric pitch angle θ_p, which is a function of propeller diameter D_p and propeller pitch H_p, and effective blade angle α_{ab} are defined by Equations (2a) and (2b). The corresponding analytical thrust, blade-drag and torque coefficients follow from Equations (3a) - (3c). Here A is aspect ratio, ε is the downwash correction, λ is the effective blade-area factor, ζ is the mean linear-speed factor, e is the Oswald factor, C_{fd} is the profile-drag coefficient, α_0 is the zero-lift angle and K_0 is the lift-curve slope. Recommendation for the parameter ranges for two blade (B_p = 2) propellers is given in [6].

$$\theta_p = tan^{-1}\left(\frac{H_p}{\pi D_p}\right) \tag{2a}$$

$$\alpha_{AB} = \varepsilon\theta_p - \alpha_0 \tag{2b}$$

$$C_{T,a} = \frac{0.25\pi^3 \lambda \zeta^2 B_p K_0 \alpha_{AB}}{\pi A + K_0} \tag{3a}$$

$$C_d = C_{fd} + \frac{\left(\frac{\pi A K_0^2}{e}\right)\alpha_{AB}^2}{(\pi A + K_0)^2} \tag{3b}$$

$$C_{M,a} = \frac{\pi^2 C_d \zeta^2 \lambda B_p^2}{8A} \tag{3c}$$

For each raw point i, the two regression ordinates are defined by Equations (4a) and (4b). The constant coefficients are then fitted through the physical origin using Equations (5a) and (5b).

$$q_{T,i} = \rho_i \left(\frac{N_i}{60}\right)^2 D_p^4 \tag{4a}$$

$$q_{M,i} = \rho_i \left(\frac{N_i}{60}\right)^2 D_p^5 \tag{4b}$$

$$\hat{C}_T = \frac{\sum q_{T,i} T_i}{\sum q_{T,i}^2} \tag{5a}$$

$$\hat{C}_M = \frac{\sum q_{M,i} M_i}{\sum q_{M,i}^2} \tag{5b}$$

With the parameters recommended from [6] for two bladed propellers: A = 5, ε = 0.85, λ = 0.75, ζ = 0.5, e = 0.83, α_0 = 0°, C_{fd} = 0.015 and K_0 = 6.11 [6]; the obtained values for our case are C_T = 0.1638 and C_M = 0.0174. The experimental fit gives C_T = 0.1506 and C_M = 0.0099. This gives significant deviations between analytical and experimental value of coefficients. Since the parameter B_p affects certain terms in equations (3a) and (3c), it is necessary to determine appropriate values for the other parameters.

Measurable parameters are changed according to case, and others analytical parameters were determined by minimizing the relative-error objective in Equation (6).

The aspect ratio A is fixed at 12.7 and it is based of used 10x5x3-in propeller; ε, λ, ζ, e and α_0 are constrained to the prescribed intervals [6]. Because λ and ζ enter only through $\lambda\zeta^2$, their individual values are not uniquely identified; the feasible pair nearest the centers of the two ranges is retained.

$$J = \left(\frac{C_{T,a}}{\widehat{C_T}} - 1\right)^2 + \left(\frac{C_{M,a}}{\widehat{C_M}} - 1\right)^2 \quad (6)$$

The solution is A = 12.7, ε = 0.918, λ = 0.812, ζ = 0.621, e = 0.80, α_0 = -0.625°, C_{fd} = 0.015 and K_0 = 6.11. It gives, C_d = 0.0363, C_{Ta} = 0.1506 and C_{Ma} = 0.0099. The experimental fit gives C_T = 0.1506 and C_M = 0.0099. Their numerical equality is a result of calibration and is not an independent validation of the geometric model.

2.3. LY-10KGF experiment and data treatment

Testing was performed on a LY-10KGF test stand Figure 3 using MET-V6 acquisition software. For the test, a single battery-ESC-motor-propeller combination was used. Between each of the four experimental measurements, the battery was fully recharged and the motor was allowed to cool to ambient temperature. The test stand specification and vendor software documentation define the measured channels. The LY-10KGF specification gives 10 kgf (98.07 N) thrust range with 0.001 kgf (0.0098 N) resolution and an error expression of 0.1% of reading plus 0.1% full scale. The corresponding torque figures are 5 Nm, 0.001 Nm resolution and 0.2% of reading plus 0.2% full scale [14,15]. Four exports contain command, optical RPM, voltage, DC current, thrust, torque, temperature and environmental data.

Figure 3. LY-10KGF test stand with GEPRC EM3115 and 10x5x3 in three blade propeller

Each of the four experimental sweeps followed a descending command sequence. The command was reduced from 100% to 10% in 10-percentage-point increments, with each level held for 5 s, and was then set to 0% to stop the system. For each sweep and command level, the samples recorded during the corresponding 5-s interval were time-averaged to obtain one representative value for each measured channel.

The original records contain ten nominal command levels, but all 90% and 100% rows are removed before calculation and plotting. These results were not used in analysis because of ESC rating and motor power limitations. At 90%, the mean branch current is 64.27 A, too close to the 65 A rating to provide a useful operating margin; at 100%, every measured branch current exceeds the rating. The retained analysis therefore contains 32 sweep-level points, corresponding to eight command levels from 10% to 80% across four sweeps. The test stand manufacturer recommends testing propulsion units producing more than 1 kgf (9.807 N) [14]. The 10÷30% points are below that level, and the 10% point has 9.7÷19.7% coefficient of variation in principal channels. The coefficient fit is restricted to 20 points from 40% to 80%. For each retained command, the arithmetic mean and sample standard deviation across four measurements are reported. The symbol ± s in the tables and figures is defined by Equation (7). s_x is standard deviation, n = 4 number of repeated measurements, x_i individual value measured during the i-th experimental run, $\bar{x}$ arithmetic mean value of all measurements.

$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} \quad (7)$$

Agreement is evaluated with the root-mean-square error RMSE and mean absolute percentage error MAPE definitions in Equations (8a) and (8b), where y_i is the experimentally measured thrust or torque value at the i-th data point, and $\hat{y}_i$ is the corresponding thrust or torque value predicted by the model. Leave-one-sweep-out cross-validation was applied. In each of the four iterations, C_T and C_M were identified from the 40÷80% data of three sweeps and evaluated on all retained points of the remaining sweep. This procedure tests whether the identified coefficients are transferable between independent experimental runs and avoids evaluating the model using the same data employed for fitting.

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \quad (8a)$$

$$MAPE = \frac{100}{n}\sum\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8b)$$

Piecewise-quadratic curves in Figure 4 are visual guides through the retained command means only. In Figure 5, both analytical curves are direct evaluations of Equations (1a) and (1b). The forward-flight map uses interpolation in measured thrust and remains entirely within the measured 40÷70% interval.

2.4. Electrical and vehicle-level calculations

The analytical motor current model uses the calibrated C_M and the interpolated optical RPM. With nominal no-load motor constant K_{V0} = 900 RPM/V, nominal no-load current I_{m0} = 2 A at nominal no-load voltage U_{m0} = 24 V and motor resistance R_m = 0.041 Ω, motor current follows Equation (9).

$$I_m = \frac{MK_{V0}U_{m0}}{9.55(U_{m0} - I_{m0}R_m)} + I_{m0} \quad (9)$$

Equations (10a) - (10d) close the motor, ESC and battery model. The ESC resistance is R_e = 0.008 Ω, battery resistance is R_b = 0.004 Ω and a conservative battery voltage of U_b = 21 V is adopted for forward-flight performance analysis. The physically admissible smaller root, $0 \le \sigma \le 1$, is used. K_E is back electromotive force constant, and U_{req} is the effective voltage required by the ESC-motor branch to sustain the prescribed rotational speed and torque, including the voltage drop across the ESC resistance.

$$K_E = \frac{U_{m0} - I_{m0}R_m}{K_{V0}U_{m0}} \quad (10a)$$

$$U_m = I_mR_m + K_EN \quad (10b)$$

$$U_{req} = U_m + I_mR_e \quad (10c)$$

$$U_{req} = \sigma[U_b - R_b(4\sigma I_m + I_{other})] \quad (10d)$$

Shaft power P_{shaft} and electrical-to-shaft efficiency η are calculated from synchronized channels by Equations (11a) and (11b). U is the measured DC voltage at the input of the tested ESC-motor branch.

$$P_{shaft} = \frac{2\pi NM}{60} \quad (11a)$$

$$\eta = \frac{P_{shaft}}{UI_e} \quad (11b)$$

For four equal branches and I_{other} = 1 A auxiliary consumption, total current I_b and constant-current endurance in minutes t_{fly} are given by Equations (12a) and (12b). Auxiliary consumption of 1 A is determined from datasheet of other electrical peripherals, and it was confirmed by measuring and rounding up to higher value. Battery capacity is C_b = 10 Ah and usable capacity f_{use}=0.8K_b; 80% of total battery capacity K_b, the constant-current endurance is:

$$I_b = 4I_e + I_{other} \quad (12a)$$

$$t_{fly} = \frac{60f_{use}C_b}{I_b} \quad (12b)$$

The ideal payload and force-balance tilt θ associated with a selected per-rotor thrust limit are given by Equations (13a) and (13b).

$$m_{pay} = \frac{4T_{lim}}{g} - m$$
$$m_{pay} = \frac{4T_{lim}}{g} - m \quad (13a)$$

$$\theta_{max} = arccos\left(\frac{mg}{4T_{lim}}\right) \quad (13b)$$

For conditional forward flight, required thrust per rotor is first evaluated from Equation (14).

$$T_{req}(\theta) = \frac{mg}{4cos\theta} \quad (14)$$

The conditional speed bound combines static drag balance, geometric pitch speed and the stricter minimum in Equations (15a)-(15c). V_D is the static drag-balance speed, V_p is the geometric pitch-speed bound and V_{lim} is their minimum. The projected area varies from 0.028 to 0.035 m² and the assumed drag coefficient from 2.05 to 1.80; these are unmeasured screening inputs based on bluff-body practice [16].

$$V_D = \sqrt{\frac{2mgtan\theta}{\rho S(\theta)C_D(\theta)}} \quad (15a)$$

$$V_p = \frac{NH_p}{60} \quad (15b)$$

$$V_{lim} = min(V_D, V_p) \quad (15c)$$

Finally, Equation (16) converts conditionally limited speed and endurance t_{fly} in minutes into conditional range R in kilometers.

$$R = \frac{V_{lim}t_{fly}60}{1000} \quad (16)$$

3. RESULTS

3.1. Static electrical and mechanical results

Electrical input power U·I exceed shaft power because part of the supplied power is dissipated within the ESC–motor–propeller system. The difference increases with load, primarily due to current-dependent losses and the accompanying reduction in battery voltage. The electrical-to-shaft efficiency increases from 40.2% at 10% command to the highest measured value of 65.5% at 70%, before decreasing to 63.8% at σ = 80%.

Table 2 summarizes the retained pooled measurements. From 10% to 80%, current, thrust, and torque increase monotonically with command. At the highest point the mean branch current is 44.919 A, safely below the 65 A reference line; the excluded 90% regime would leave less than 1 A margin in the highest sweep.

The corresponding current and power trends are shown in Figure 4. Electrical input power U·I exceed shaft power because part of the supplied power is dissipated within the ESC–motor–propeller system. The difference increases

with load, primarily due to current-dependent losses and the accompanying reduction in battery voltage. The electrical-to-shaft efficiency increases from 40.2% at 10% command to the highest measured value of 65.5% at 70%, before decreasing to 63.8% at σ = 80%.

Table 2. LY-10KGF measurements (mean ± sample SD)

σ [%]	N [RPM]	I [A]	T [g]	M [Nm]	η [%]
10	2246±218	0.43±0.08	88±17	0.018±0.003	40.2
20	4833±53	2.81±0.06	446±10	0.080±0.001	58.2
30	6613±49	6.70±0.09	875±15	0.147±0.001	61.8
40	7956±63	11.81±0.07	1285±12	0.213±0.002	61.5
50	9109±67	17.34±0.14	1683±21	0.276±0.003	62.6
60	10128±41	23.71±0.07	2093±9	0.345±0.002	64.1
70	11148±1	31.82±0.59	2587±35	0.425±0.008	65.5
80	12219±97	44.92±0.14	3195±9	0.526±0.003	63.8

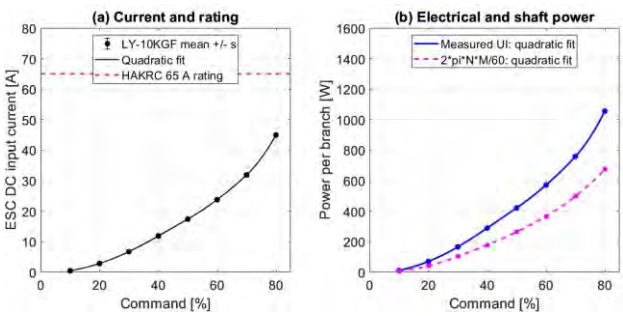


Figure 4. Retained branch current and 65 A ESC rating (a), and electrical and shaft power (b)

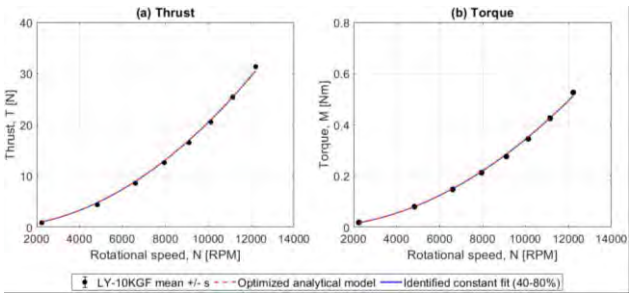


Figure 5. Static thrust (a) and torque (b) versus optical RPM

3.2. Calibrated and identified coefficients

Identification using 20 raw data points from 40% to 80% gives C_T=0.1506 and C_M=0.0099. The constrained geometric optimization produces the same values to plotting precision. **Table 3** reports the agreement metrics, and Figure 5 compares both models with pooled measurements.

Table 3. Agreement in the 40÷80% identification region

Model	C_T	C_M	T MAPE [%]	M MAPE [%]
Analytic geometry	0.1506	0.0099	2.274	2.331
Identified constant	0.1506	0.0099	2.274	2.331

The optimized geometric and identified constant curves overlap because their coefficients are numerically equal.

When evaluated against the same 20 measurements from the 40÷80% command range that were used for coefficient identification, the fitted model gives RMSE 0.5747 N for thrust and 0.00998 Nm for torque; the corresponding centered R^2 values are 0.9924 and 0.9918. Leave-one-sweep-out validation increases MAPE only slightly, from 2.274% to 2.326% for thrust and from 2.331% to 2.419% for torque. This small increase indicates that a single constant pair of C_T and C_M provides a repeatable approximation for the tested 10×5×3 propeller under static conditions within the 40÷80% command range. However, the results do not justify applying the same coefficients outside this operating range or to another propeller.

3.3. Vehicle-level analysis

Per-sweep interpolation to 1350 g per rotor gives hover at 41.65±0.32% command, 8146±91 RPM and 12.715±0.134 A per branch. Equations (12a) and (12b) give 51.86±0.54 A total and t_{fly}=9.26±0.10 min. At 70%, measured branch current is 31.82 A, total current is 128.29 A and t_{fly}=3.74 min. Mean thrust of 2587.3 g per rotor gives necessary vertical trust for steady horizontal flight at 58.55° pitch angle. Table 4 summarizes these results.

Table 4. Conditional vehicle-level results

Quantity	Conditional result
Hover	41.65%; 8146 RPM; 9.26 min
70% point	31.82 A/branch; 3.74 min
Tilt angle at σ = 70%	58.55°
Measured-current max range at θ = 12.34 °	9.363 km; 17.44 m/s; 8.95 min
Analytical-current max range θ = 12.34 °	10.338 km; 17.44 m/s; 9.88 min
Maximum conditional speed bound at σ = 70%	23.6 m/s at 58.55°

The corresponding speed and range maps are presented in Figure 6. In Figure 6a. Up to 12.344°, V_D is active; beyond the crossing, V_p is active. In Figure 6b, the blue curve uses the analytical BLDC/ESC current model with experimentally calibrated propeller loading, whereas the magenta curve directly scales the measured LY-10KGF branch current.

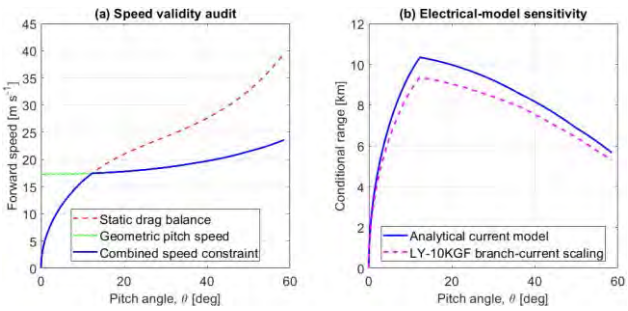


Figure 6. Conditional forward-flight analysis: speed constraints (a) and range sensitivity to the current model (b)

Both range curves peak at $\sigma = 42.5\%$, $\theta = 12.34°$ and $V_{lim} = 17.44$ m/s. When the measured motor current is used, the estimated range is 9.363 km with $t_{fly} = 8.95$ min, while the analytical predicted current model gives 10.348 km of range with $t_{fly} = 9.88$ min, 10.42% higher. The apparent sharp decrease is a slope change at the min (V_D,V_p) constraint switch: thereafter speed grows slowly, while required thrust, current and endurance penalty continue to rise.

4. DISCUSSION

The final analytical and identified coefficients are intentionally equal. The geometric parameters were optimized against the experimental C_T and C_M, so the overlapping curves in Figure 5 demonstrate successful calibration rather than an independent theoretical prediction. The analytical formulation remains useful because it shows how the aspect ratio, blade-area factor, mean-speed factor, downwash correction and drag assumptions influence the coefficients. However, λ and ζ cannot be interpreted separately from this data set because only their product $\lambda\zeta^2$ is observable. The next step is to test other three-bladed propellers with suitable motors in order to validate or refine the selected parameters.

The exclusion of the 90% and 100% command levels is justified on both electrical and methodological grounds. The 90% mean does not exceed 65 A, but individual runs approach the rating and leave no practical current or control reserve; every 100% run exceeds the rating. Removing both regimes before analysis avoids fitting a local static model in a region dominated by component limits and heat accumulated during preceding runs. The highest retained 80% point remains within the controller rating.

The corresponding current and power trends are shown in Figure 4. The electrical input power U·I exceeds the shaft power because part of the supplied power is dissipated through losses in the ESC and motor. The absolute difference between electrical input power and shaft power increases with loading, which is consistent with the growing influence of current-dependent losses. When efficiency is evaluated separately for each experimental measurement and then averaged across the four repetitions, it rises overall from 40.2% at 10% command to the highest measured value of 65.5% at 70%, before decreasing to 63.8% at 80%.The low-command value is particularly sensitive to measurement uncertainty because both torque and input power are small. Its generally lower value reflects the greater relative contribution of no-load and weakly load-dependent losses, whereas the decrease at 80% is consistent with increasing current-dependent losses. The reported quantity is the electrical-to-shaft efficiency of the tested ESC–motor branch under static propeller loading and does not include the aerodynamic conversion efficiency of the propeller.

Directly measured current is the stronger basis for endurance because it avoids propagating any residual torque-model error through the motor equations. The 10.42% separation between the two Figure 6b curves quantify model sensitivity. The total value 128.29 A at 70% is the sum of four branches and 1 A auxiliary load; it must not be compared with the 65 A rating of one ESC.

Figure 6 remains a conditional analysis. The surface area and drag coefficients were not measured, V_p omits slip and inflow, static data do not represent oblique flow, and the calculation assumes four equally loaded rotors and constant usable battery capacity.

Future validation should combine bidirectional airspeed, attitude, RPM, voltage, current and temperature measurements with controlled battery state. Raw synchronized samples, tachometer redundancy, calibration certificates and repeated zero checks would permit a complete uncertainty budget rather than descriptive between-sweep repeatability alone.

5. CONCLUSION

Four LY-10KGF measurements provide 32 retained points from 10% to 80%. A 40-80% least-squares fit gives $C_T = 0.1506$ and $C_M = 0.0099$. Constrained calibration with $A = 12.7$, $\varepsilon = 0.918$, $\lambda = 0.812$, $\zeta = 0.620961$, $e = 0.80$ and $\alpha_0 = -0.625°$ reproduces both coefficients. Leave-one-run-out MAPE is 2.326% for thrust and 2.419% for torque.

The results indicate that the GEPRC EM3115–HAKRC–10×5×3 combination is adequately, but not optimally, matched within the retained operating range. The 65 A ESC is adequate up to the highest retained 80% command level; however, the limited current margin at 90% and the rating exceedance at 100% indicate that it is not suitable for unrestricted operation near full command. An ESC with a higher continuous-current rating would provide greater transient-current and thermal margin and reduce the risk of controller overload. A broader continuous operating range could also be achieved by reducing the propeller load through a smaller pitch, diameter or blade number, or by selecting a motor with a lower K_V or a higher continuous-current and thermal rating. These changes must be evaluated as a complete motor–ESC–propeller system, since a higher-rated ESC or motor alone would not reduce the required current or necessarily improve efficiency.

Measured current gives hover at 8146 RPM and 41.65% command with t_{fly}=9.26 min. At 70%, t_{fly} falls to 3.741 min, while the conditional force-balance tilt reaches 58.55°. The maximum conditional range is 9.363 km when calculated from measured branch current and 10.338 km when calculated from the analytical current model. These values

support component selection and transparent design comparison, but not flight certification or validation.

Acknowledgment

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03-34/2026-03/200325).

References

[1] ANTONIĆ, V., TRIFKOVIĆ, M., MOLOVIĆ, V., MILENKOVIĆ-BABIĆ, M.: *Small Fixed-Wing UAV Precision Aerial Drop Capability Development*, Scientific Technical Review, 74(2) (2024) pp 65-70

[2] U.S. ARMY: RQ-11B Raven Small Unmanned Aircraft System, army.mil, 4 November 2014.

[3] U.S. ARMY CENTER FOR ARMY LESSONS LEARNED: Integrating sUAS and Loitering Munitions into Company and Below, No. 26-1148, April 2026.

[4] GILL, R., D'ANDREA, R.: *Propeller thrust and drag in forward flight*, IEEE CCTA, 2017, pp. 73-79. doi:10.1109/CCTA.2017.8062443.

[5] SHI, D., DAI, X., ZHANG, X., QUAN, Q.: *A practical performance evaluation method for electric multicopters*, IEEE/ASME Transactions on Mechatronics, 22(3) (2017), pp. 1337-1348. doi:10.1109/TMECH.2017.2675913.

[6] QUAN, Q.: Introduction to Multicopter Design and Control, Springer, Singapore, 2017. doi:10.1007/978-981-10-3382-7.

[7] HWANG, M.-H., CHA, H.-R., JUNG, S.Y.: Practical endurance estimation for minimizing energy consumption of multirotor unmanned aerial vehicles, Energies, 11(9) (2018), 2221.

[8] THEYS, B., DIMITRIADIS, G., HENDRICK, P., DE SCHUTTER, J.: *Experimental and numerical study of mini-UAV propeller performance in oblique flow*, Journal of Aircraft, 54(3) (2017), pp. 1076-1084.

[9] THEYS, B., DE SCHUTTER, J.: *Forward flight tests of a quadcopter UAV with various spherical body diameters*, International Journal of Micro Air Vehicles, 12 (2020), pp. 1-8.

[10] YE, J., WANG, J., SONG, T., et al.: *Propulsion optimization of a quadcopter in forward state*, Aerospace Science and Technology, 113 (2021), 106703.

[11] POLLET, F., DELBECQ, S., BUDINGER, M., MOSCHETTA, J.-M.: *Design optimization of multirotor drones in forward flight*, ICAS 2021, HAL hal-03832135.

[12] GEPRC: EM3115 900KV Motor - Official Product Specification , https://geprc.com/product/geprc-em3115-900kv-motor/, accessed 12 July 2026.

[13] HAKRC: 65A ESC - Official Product Specification, https://www.hakrc.com/HAKRC-4234-6S-65A-Airplane-Monobloc-ESC.html, accessed 12 July 2026.

[14] WING FLYING TECHNOLOGY: LY-10KGF Motor and Propeller Test System - Specification, 2021.

[15] WING FLYING TECHNOLOGY: MET-V5 Software User Manual, 2021.

[16] HOERNER, S.F.: Fluid-Dynamic Drag, published by the author, Midland Park, NJ, 1965.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SIMULATING SHOCK-WAVE/BOUNDARY-LAYER INTERACTIONS

JELENA SVORCAN
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, jsvorcan@mas.bg.ac.rs

DRAGOLJUB TANOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, dtanovic@mas.bg.ac.rs

MILICA IVANOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, mivanovic@mas.bg.ac.rs

Abstract: *At transonic and supersonic operating regimes, shock and expansion waves form and interact with boundary layer forming along the walls of the body moving through the fluid. In such cases, both compressibility and viscosity effects must be considered in order to obtain valid estimations of the pressure field and resulting aerodynamic forces, as well as some specific flow features such as separation point, wake shape, etc. The paper presents several numerical studies of the relatively standard transonic/supersonic flow cases (around the airfoil, inside a duct where reflected waves also appear, etc). Its main goals are to test different numerical set-ups and provide a quantification of the possible accuracy and reliability of the inspected numerical models that can be achieved when shock-wave/boundary-layer interactions might be of importance. Different computational meshes, turbulence models, solvers, etc. are tried and compared. The most significant conclusions and recommendations are provided in the end.*

Keywords: *shock-wave, boundary-layer, interaction, CFD.*

1. INTRODUCTION

Both shock-waves and boundary-layers appearing along the walls are still regarded today as distinct fluid flow phenomena, insufficiently resolved but of great importance for aerospace applications since they directly affect the aerodynamic forces and consequently aerodynamic efficiency [1-7]. They are particularly interesting to study when combined together in viscous, compressible flows (highly-subsonic, transonic and supersonic) since they may produce significant drag increase leading to severe separation zones as well as flow instabilities and difficulties with control, directly affecting flight envelope and safety. For all these reasons, the topic of shock-wave/boundary-layer interaction (SWBLI) is still very important and actively investigated, particularly the ways to lessen its potentially negative effects [4,5].

The most employed research approaches for studying SWBLI are numerical and experimental [1-7]. Ideally, computational fluid dynamics (CFD) methods usually require very fine meshes and advanced turbulence modelling. Due to the great computational cost and small temporal and spatial scales required, unsteady Reynolds-averaged Navier-Stokes (URANS) equations are still very much employed [5,6,7], even though different scale-resolving methods (such as large eddy simulation, LES, and detached eddy simulation, DES) are increasingly tried [1-3,7]. In both cases, experimental validation is preferred since very complex flow phenomena is investigated. Both quantitative (pressure and force measurements, particle image velocimetry, etc.) and quantitative methods (e.g. Schlieren photography) are common [4].

In the past decades, CFD techniques have been utilized very much. They have now reached the level of great accuracy in addition to being much more user-friendly than before (e.g. mesh generation process is much facilitated as is the solving of systems of equations). Available computational power is also much increased. However, CFD methods still perform the best in the cases of attached flows and nominal operating conditions, and additional validation in cases involving any kind of flow phenomena is necessary. Among the recognized flow phenomena, the SWBLI is certainly one of the most important and interesting since most of the governing mechanisms are still not explained in detail. Even though LES and DES technics are recommended today in such flow cases, the extents of (U)RANS applicability should also be determined, particularly when combined with refined numerical grids and spatial geometric models (planar analysis by standard turbulence models proved inadequate).

This paper is organized in the following manner. The flow around the airfoil is studied in chapter 2. The flow in the channel with a shock generator is described in chapter 3. The flow around a bluff body in a channel is presented in chapter 4. All these separate sections contain details regarding the employed computational models, meshes, numerical set-ups and obtained results. Finally, some of the most important conclusions are offered in the last chapter.

Corresponding author: J. Svorcan;
E-mail: jsvorcan@mas.bg.ac.rs
DOI: 10.5937/OTEH26004S

2. EXAMPLE 1

Example 1 studies the transonic flow around a supercritical airfoil.

2.1. Expected flow physics (sonic bubble)

Highly-subsonic flows, approaching transonic, usually involve a sonic bubble on the upper surface of the airfoil ending in a shock-wave, and affecting the subsequent flow separation. Additional complexity may be transonic buffet, a transient flow phenomenon, caused by the dynamic interaction of shock and boundary layer. It has often been studied, and some representative cases have been established. Type II transonic buffet where the shock oscillatory moves (back and forth) is particularly interesting [1]. It depends on the combination of angle-of-attack α, Mach number M and Reynolds number Re (and even the type of flow transition), and involves an oscillating sonic bubble on the airfoil suction side.

2.2. Airfoil geometry, operating conditions

This study seeks for the simplest computational setup capable of successfully simulating highly-subsonic flow (that might also include a transonic buffet) around the ONERA OAT15A supercritical airfoil at the different Mach numbers, M = 0.715 and M = 0.730, where transient flow is not and is respectively expected. In both cases, Reynolds number is Re = 3,000,000 (computed with respect to the chord). This airfoil, designed to save fuel, with a flattened top surface and 12.5% maximal relative thickness, is widely used in aerospace research and has a recognized "shaking" effect.

2.3. Computational domain, mesh and numerical set-up

A relatively simple, standard computational domain and meshes are employed. The airfoil OAT15A, generated from the publicly available coordinates [8], is surrounded by a circular domain (extending 10 chord lengths c in radial direction) and extruded by 20% c in the third, spanwise direction. Initially, 2-dimensional flow simulations were tried, but they appeared completely incapable of capturing the flow nuances of interest even though planar buffet is the main focus here (and not a spatial one appearing on a finite wing). Fortunately, even a slight extrusion of the planar computational domain, and quasi-spatial consideration of the flow phenomena, greatly improves the accuracy and stability of the numerical study.

Hybrid, unstructured computational grids were formed, numbering several million cells (a detail is illustrated in Fig. 1). First layer thickness corresponded to the dimensionless wall distance $y^+ \approx 5$, whereas meshes were additionally refined along the airfoil. In the second iteration, adapted numerical grids were used (Fig. 1, bottom).

Flow simulations were performed using ANSYS Fluent where the governing flow equations are solved by finite volume method [9]. Flow is steady or transient, compressible and turbulent. Turbulence was resolved either by k-ω SST turbulence model (as more computationally efficient) or by wall-modeled large eddy simulation (WMLES) (since this relatively recent model should be capable of resolving a large portion of turbulent motion). Air is assumed as the ideal gas. Pressure far-filed boundary conditions were assumed far away from the airfoil (static pressure p and static temperature T corresponding to total values p_o = 101325 Pa and T_o = 300 K, respectively, Mach number and flow direction corresponding to the angle-of-attack α = 3.5°), while the two side faces form a periodic boundary, and airfoil walls are no-slip. Pressure-based coupled solver was employed, and all spatial discretizations are 2nd order. Flow computations are performed until reaching the (quasi-) convergence of the aerodynamic force coefficients.

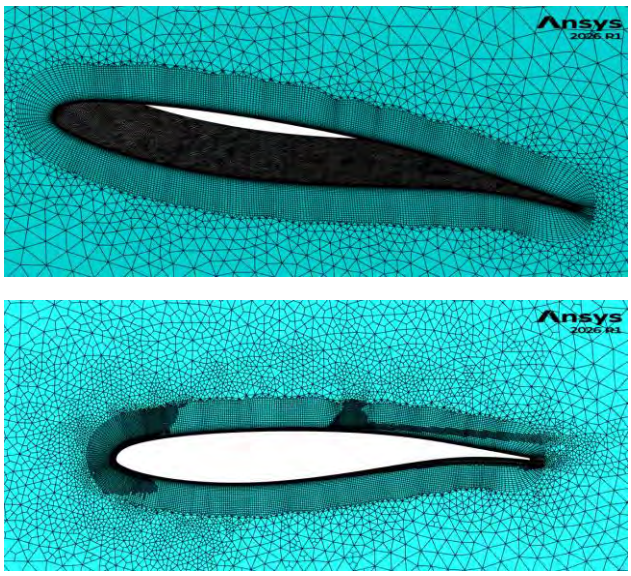
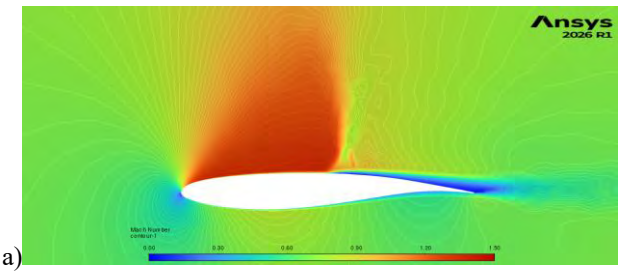


Figure 1. Details of the computational meshes around the airfoil OAT15A, original (top) and adapted (bottom)

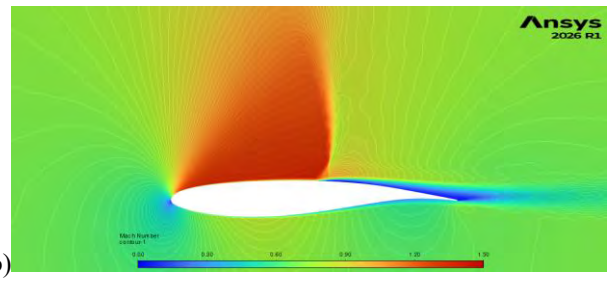
2.4. Results and discussion

Computed flow visualizations are illustrated in Fig. 2 (courser vs. finer mesh and k-ω SST vs. WMLES).

a)



b)



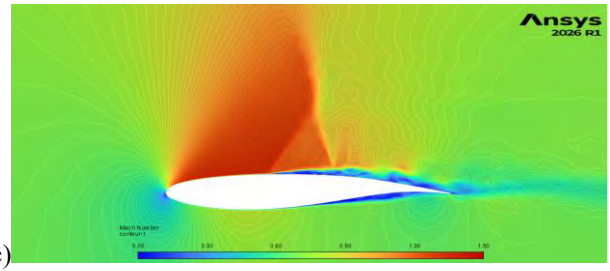
c) 

Figure 2. Instantaneous contours of Mach number around the airfoil OAT15A at M = 0.73, a) k-ω SST, coarser mesh, b) k-ω SST, adapted mesh, c) WMLES, adapted mesh (during buffet)

Overall, by using *k-ω* SST turbulence model, it is possible to obtain the main, averaged flow features. By additionally refining the mesh, the contours of shock-waves can be precisely captured. However, to capture the expected oscillations of the sonic bubble and shock-wave it is necessary to employ one of the scale-resolving turbulence models (here WMLES). Figure 2c illustrates the instantaneous contours of Mach number near its most upstream position, when boundary layer is the longest, and lift coefficient approaches its minimum. The features of the resulting λ-wave are also clearly delineated. In general, the shock-wave moves between approximately 47% to 59% chord length, when moving from its most upstream to most downstream location, respectively [1]. The oscillating shock-wave directly affects the length of the separation zone appearing on the upper airfoil surface, and induces the variability of aerodynamic coefficients.

Lift C_L and drag C_D force coefficient variability in time is depicted in Fig. 3. Mean values of lift coefficient and Strouhal number are close to the experimentally documented ones ($C_{L,\mathrm{exp}}$ = 0.95±0.15 and St ≈ 0.06, respectively [1]). However, for a more precise description of transient behavior, more refined spatial- and time-scales are required.

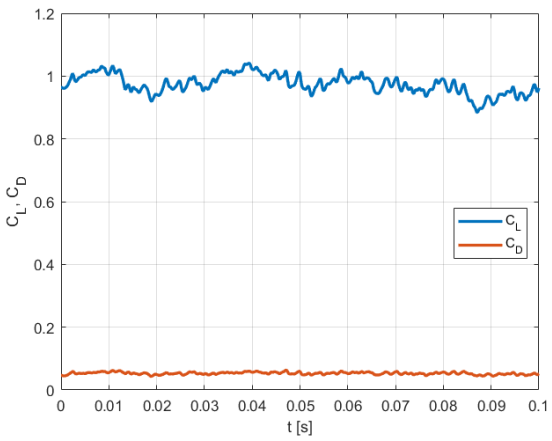


Figure 3. Oscillating aerodynamic coefficients (lift and drag) during buffet

Steady fluid flow obtained at M = 0.715 is visualized in Fig. 4 by Mach number contours. Some differences to the unsteady flow case illustrated in Fig. 2 are apparent. Most distinctly, boundary layer is thinner and λ-wave is less pronounced which leads to improved aerodynamic performance (increased lift and decreased drag coefficient).

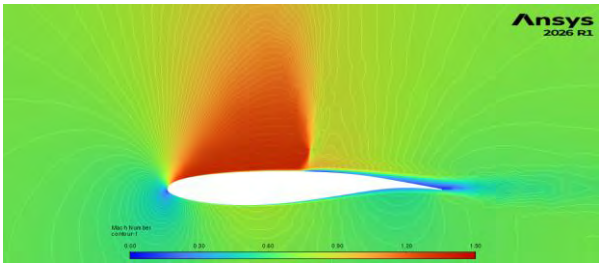


Figure 4. Steady flow at M = 0.715 computed by *k-ω* SST turbulence model

3. EXAMPLE 2

Example 2 studies the channel flow with shock generator and reflected waves.

3.1. Expected flow physics (reflected waves)

Supersonic flow through channels of variable geometry or with shock generators (like steps, wedges, etc.) usually involve a complex pattern of waves (starting from the initial shock wave that is then reflected from the opposite wall, as well as the interference of reflected waves). Additional simulating challenges stem from the interference of waves with the boundary layer that lead to recirculating zones, thickening of the boundary layer and final separation.

3.2. Geometry of the channel and wedge, operating conditions

The current study explores a simple channel flow with a 20° wedge (10 mm in length) positioned at the first third of the channel upper wall [4,5]. Mach number of the flow is M = 2.95, whereas the total length of the channel is 130 mm.

3.3. Computational domain, mesh and numerical set-up

Computational domain corresponds to the previously mentioned channel geometry.

Again, the computational mesh is hybrid (including thin layers of cells along the walls corresponding to the dimensionless wall distance $y^+ \approx 1$), unstructured, as simple and uniform as possible, initially numbering approximately 800,000 cells. Later, in the computation process, this initial mesh was adapted in the zones of increased density gradients, leading to a total number of approximately 2 million control volumes. The differences between the two meshes can be observed in Fig. 5.

In this case, due to the larger undisturbed flow Mach number (in comparison to the transonic flow around the airfoil), density-based solver was employed. Again, the flow is steady, viscous (turbulent, resolved by *k-ω* SST turbulence model) and compressible (with the assumption of ideal gas). Outlet boundary conditions are now more convoluted, i.e. harder to accurately define (due to the

apparition of local losses). Pressure outlet condition was assumed with the same value of static pressure as defined along the inlet, and with the careful monitor of the total mass flux (entering and exiting the domain). Since convergence might be difficult in cases involving complex shock-wave interaction, Courant-Friedrichs-Lewy (CFL) number was reduced to 0.5, together with all the relaxation factors. Simulations were performed until reaching the converged values of outlet mass flow rate and aerodynamic coefficients along the upper wall (where the wedge is located).

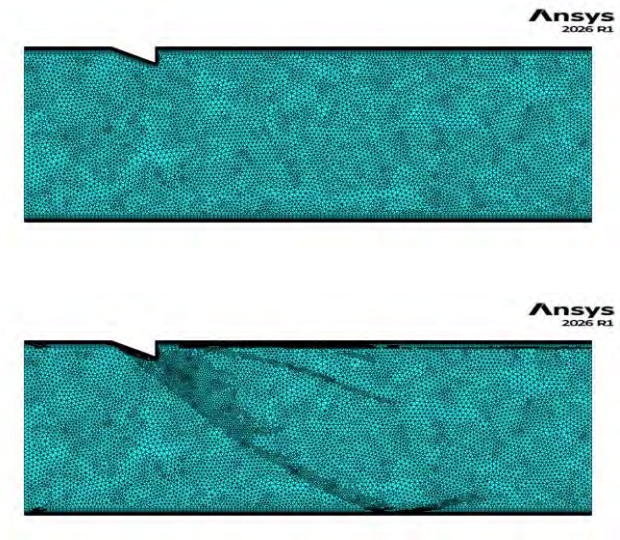


Figure 5. Differences of the initial and adapted meshes in the channel with the shock generator

3.4. Results and discussion

The obtained Mach number contours are depicted in Fig. 6.

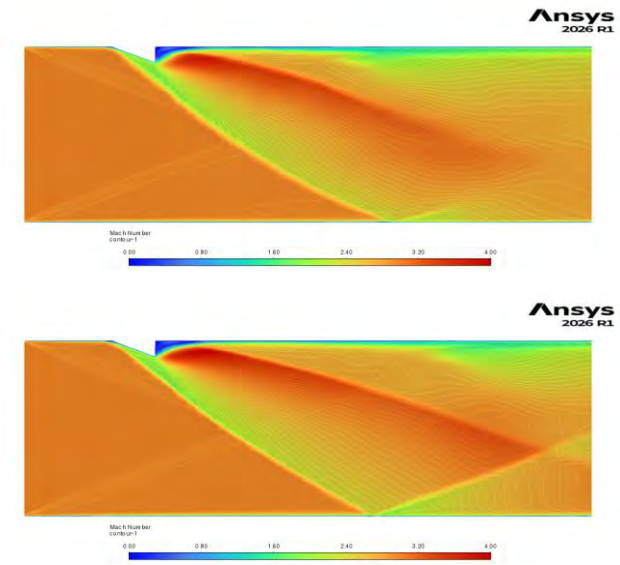


Figure 6. Contours of Mach number inside the channel (original vs. adapted grid)

It can immediately be observed that a much clearer pattern of shock waves can be obtained on the finer (locally adapted) computational grid (again by using *k-ω* SST turbulence model). The original oblique shock, induced by the wedge, forms and then reflects from the lower wall producing a small recirculating zone at approximately x/l = 0.59. At the same time, due to the wake formed at the bottom of the wedge, expansion waves are also noticeable, as well as the accompanying oblique shock, and their subsequent interaction with the reflected wave. Wake behind the wedge narrows fast, but the boundary layer along the upper wall thickens. These flow visualizations are in line with the results of other studies [4,5]. Since the resulting, averaged shock-wave pattern was the main goal here, the *k-ω* turbulence model proved sufficient. For a more accurate and detailed resolution of the wedge wake, a scale-resolving model should be employed.

4. EXAMPLE 3

Example 3 studies the flow around a bluff body with massive separation.

4.1. Expected flow physics (wide turbulent wake)

Modern high-speed, energy-efficient train transportation requires a careful examination of the (choked) fluid flow surrounding a bluff body in a channel where SWBLI is also present. The axisymmetric flow initially accelerates around the body, leading to the shock-wave appearing in the rear part that then interacts with the boundary-layer and results in massive flow separation aft of the bluff body [7].

4.2. Geometry of the bluff body, operating conditions

Geometry of the pod corresponds to the previous studies [7]. It is an axisymmetric body (with the total length of 5 m and maximal radius 0.49 m) comprising ellipsoidal fore and aft sections, with the cylindrical, tangentially-continuing medium part. The pod is located in a cylindrical channel with the blockage ratio 0.2.

4.3. Computational domain, mesh and numerical set-up

One twelfth of the original geometry is modelled, since axisymmetric flow is prevailingly expected (apart from the wake). The channel extends 2 and 5 pod lengths fore and aft of the model, respectively.

Again, the initial computational grid is hybrid unstructured, with the limited cell size across the computational domain, local refinement of the cells representing the pod walls, and layers of thin prismatic cells along both of the wall surfaces (pod and channel walls). Original mesh comprises approximately 320,000 cells (corresponding to 3,800,000 cells for the entire domain), while the adapted one grew to nearly 900,000 cells, Fig. 7.

Since the investigated inlet velocity corresponds to M = 0.625 pressure-based solver is employed (numerical set-

up is similar to the one employed in Example 1). The remaining boundary conditions are as follows: static pressure p = 10,000 Pa, static temperature T = 300 K, turbulence intensity 1% and dynamic viscosity ratio μ_t/μ = 10. Turbulent flow was resolved by both k-ω SST turbulence model and WMLES.

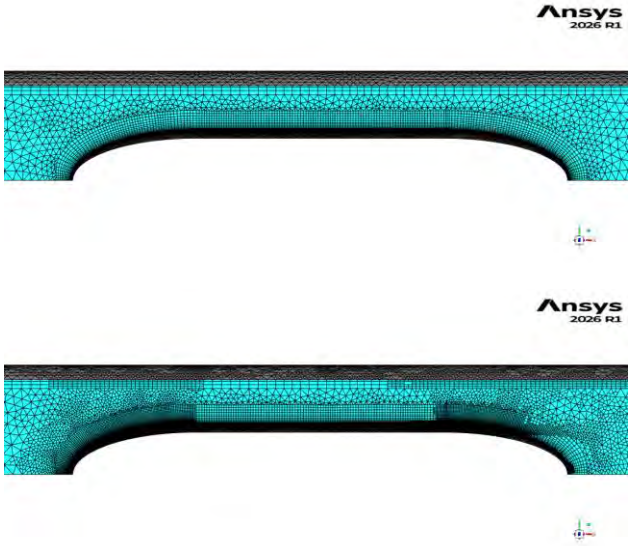
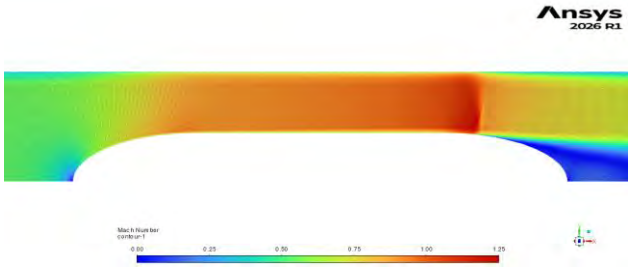


Figure 7. Details of the computational meshes around the bluff body

4.4. Results and discussion

According to the available data [7], the expected pod drag coefficient is C_D = 0.8 while the mass flow rate is 7.5 kg/s. Here, the computed mass flow rate closely matches the expected value, whereas the drag coefficient estimated by k-ω SST turbulence model and WMLES is 0.6 and 0.8 (averaged), respectively. Qualitative flow visualization is represented in Fig. 8. Even though steady vs. instantaneous flows are compared, it is apparent that URANS underestimates the size of the wake (attenuates the complex vortical structures appearing along the aft section of the pod). It is also noticeable that both the pod geometry and the shock-wave position directly affect boundary-layers and their separation on both walls, corresponding to both the body/pod and the surrounding channel. Again, when more complex (scale-resolving) turbulence model is employed, the shock-wave moves backwards and strongly affects the wake. However, it must be noted that for the adequate representation of the wake, very fine meshes must be employed.



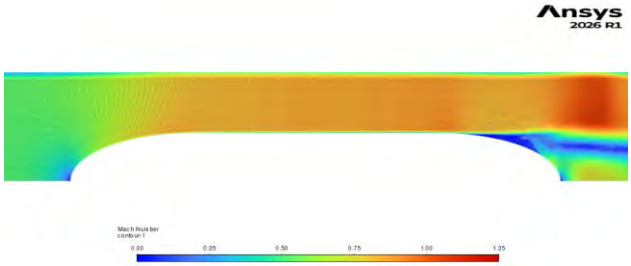


Figure 8. Instantaneous contours of Mach number around the bluff body, URANS (top) and WMLES (bottom)

5. CONCLUSION

This study numerically examined several typical flow cases involving the complex interaction between shock-waves and (turbulent) boundary-layers that are extremely important in aerospace engineering (and can be correlated to flows over isolated wings, inside supersonic inlets, and around isolated fuselages). In the future, studying flows around the full aircraft configuration might also be interesting, as well as extending the study to include the aspects of aeroacoustics.

Whereas some standard numerical setups were employed (prevailingly URANS), representative flow visualizations (such as Mach number and density contours indicating shock patterns, and separation zones pointing to thickened boundary-layers) that were obtained satisfactorily match the results of other authors. However, employing finer numerical grids together with more complex turbulence models should certainly lead to additional increase in the accuracy and reliability of numerical results, particularly when dynamic effects are the research focus, and that course shall be explored further in the future. For now, it is highly recommended to employ at least 2.5D analysis in similar cases, and locally refined meshes, particularly in the zone of expected SWBLI (but also zones fore and aft of it) as well as the wake. If any kind of massive separation is present, scale-resolving turbulence models and very fine temporal- and spatial scales should be employed.

Acknowledgement

This research work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia through contract No. 451-03-34/2026-03/200105 from February 5th 2026.

References

[1] FUKAMI,K., IWATANI,Y., MAEJIMA,S., ASADA,H., KAWAI,S.: *Compact representation of transonic airfoil buffet flows with observable-augmented machine learning*, Journal of Fluid Mechanics, 1021 (2025) A39.

[2] FUKUSHIMA,Y., KAWAI,S.: *Wall-modeled large-eddy simulation of transonic airfoil buffet at high Reynolds number*, AIAA Journal, 56(6) (2018) 2372-2388.

[3] GAO,S., DONG,H., WU,J., LIU,D., LI,G.,

YANG,L., KONG,W., GONG,M., ZHAO,Y., GU,Y.: *Study on Reynolds Number Effects on Transonic Buffet of Supercritical Airfoils with Transition Considerations*, Aerospace, 13(1) (2026) 1.

[4] DU,Z.-B., SHEN,C.-B., HUANG,W., XU,X.-P., WEI,F., YAN,X.-T., YANG,W.-Y.: *Experimental exploration on the performance of the secondary recirculation jet for the shock wave/boundary layer interaction and mixing augment in supersonic crossflow*, Aerospace Science and Technology, 176 (Part A) (2026) 112099.

[5] ZENG,F., LUO,Q., GUO,G., HE,G.: *Optimization of pressure-driven bleed–blow loop for controlling shock wave/turbulent boundary layer interactions in supersonic flows*, Acta Astronautica, 245 (2026) 740-758.

[6] CASTILLO GOMEZ,L.E., GROSS,A.: *Numerical investigation of swept turbulent shockwave boundary layer interactions*, Aerospace Science and Technology, 162 (2025) 110185.

[7] GEIBEN,B., HAVERMANN,M., HALE,E., BIL,C.: *Application of hybrid-temporal LES to shock-induced boundary layer separation in hyperloop flows*, CEAS Aeronautical Journal, 17 (2026) 1237-1258.

[8] https://www.aiaa-dpw.org/geometry.html

[9] https://www.ansys.com/

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

HIGH-PRESSURE JET SUPPLY SYSTEM FOR POWERED WIND TUNNEL MODELS: CALIBRATION AND PERFORMANCE

BILJANA ILIĆ
Military Technical Institute, Belgrade, biljana.ilic@icloud.com

MARIJA SAMARDŽIĆ
Military Technical Institute, Belgrade, marija.samardzic.vti@gmail.com

BOŽIDAR SIMOVIĆ
Military Technical Institute, Belgrade, simovic.bozidar@gmail.com

ISIDORA STANKOVIĆ
Military Technical Institute, Belgrade, stank151.dora@gmail.com

OGNJEN BABIĆ
Military Technical Institute, Belgrade, ognjen813@gmail.com

GORAN OCOKOLJIĆ
Military Technical Institute, Belgrade, ocokoljic.goran@gmail.com

Abstract: *This paper presents the calibration and performance characterization of a high-pressure jet supply system developed for experimental investigations of powered wind tunnel models. The system supplies high-pressure air through a custom flow-through internal strain-gage balance to interchangeable sonic and supersonic nozzles mounted on the wind tunnel model. Tests conducted under no-main-flow conditions evaluated the transient response of the system, quantified pressure losses along the internal flow path, determined actual jet characteristics from direct thrust measurements, and assessed nozzle operating regimes. The obtained results provide essential baseline data on the internal flow behavior and jet performance. These data will support reliable interpretation and analysis of phenomena during subsequent wind tunnel tests with powered models.*

Keywords: *powered wind tunnel model; jet thrust; flow-through balance; pressure losses; reaction control system.*

1. INTRODUCTION

Powered wind tunnel testing plays a critical role in the development of aerospace vehicles, particularly for configurations relying on reaction control systems (RCS), or other propulsion-based control mechanisms. These tests enable the simulation of jet interactions with external flow, providing essential data on aerodynamic interference, thrust vectoring effectiveness, and overall vehicle stability and control under representative flight conditions [1]. Accurate replication of jet exhaust or control thrusters is challenging due to the need for precise matching of jet pressure ratios, mass flow rates, and plume characteristics while integrating high-pressure supply systems within constrained model geometries without compromising force and moment measurements [2,3].

Traditional approaches to jet simulation in wind tunnels often employ cold or heated gas routed through the model support system [2]. Previous studies have highlighted the importance of detailed calibration of jet supply systems to quantify pressure losses, transient response, and actual delivered jet performance. For instance, investigations into RCS jet interactions for planetary entry vehicles, such as the Mars Science Laboratory, underscored the need for high-fidelity simulation to reduce uncertainties in control authority and validate computational predictions [4]. Flow-through balances are utilized to accommodate the high-pressure air supply while simultaneously measuring the resulting aerodynamic loads and jet thrust [5]. Recent advancements in flow-through balance designs have enabled more accurate isolation of jet thrust from external aerodynamic forces in powered wind tunnel testing [6].

At the VTI, ongoing research into side jet interactions on slender missile configurations has necessitated the development of a dedicated high-pressure jet supply system. This system integrates a custom flow-through internal strain-gauge balance capable of delivering high-pressure air to interchangeable sonic and supersonic nozzles mounted on the wind tunnel model [7]. The present paper focuses on the calibration and performance of this high-pressure jet supply system. Key objectives include evaluating the transient response, quantifying pressure losses along the internal flow path, and determining actual jet characteristics via direct thrust measurements. These baseline data are vital for reliable interpretation of powered wind tunnel test results.

Corresponding author: B. Ilić;
E-mail: biljana.ilic@icloud.com
DOI: 10.5937/OTEH26005I

2. EXPERIMENTAL SETUP

The details of the test facility, test model, high-pressure jet supply system and the associated measurement techniques are given in the following subsections.

2.1. Test Facility

The test facility is the VTI Belgrade T-38 wind tunnel [8]. The T-38 is a large-scale blowdown wind tunnel designed to fulfill testing requirements across a Mach number range of 0.2 to 4.0. A Reynolds number of up to 100×10^6 per meter is achieved via a controllable stagnation pressure range of 1.2 bar to 15 bar [9–11]. The 1.5 m × 1.5 m test section features a 2-DOF model support system, which permits model movement within ±20º in pitch [12] and ±720 º in roll [13].

The flow is established by discharging compressed air from 2600 m^3 storage tanks, pressurized up to 20 bar, into the atmosphere, yielding a maximum run duration of 60 s. A solid-wall test section is used for subsonic and supersonic testing, where the Mach number is set by a second throat and a flexible nozzle [14], respectively. For transonic testing, a variable-porosity test section [15] with an ejector-assisted blow-off system for active Mach number control is inserted into the wind tunnel circuit.

2.2. Test Model

The generic missile model of VTI [16] is modified to accommodate a high-pressure supply and measurement system for powered wind tunnel testing. It consists of the ogive nose cone, a cylindrical fuselage, and four wrap-around fins. The model has a nominal body diameter of D = 64 mm and a slenderness ratio of L/D = 21.6. The cross-sectional view of the model mounted on the model support sting is shown in Figure 1.

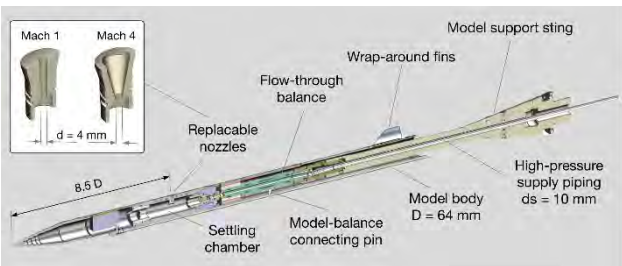


Figure 1 Powered wind tunnel model

The forward part of the model houses the settling chamber which is equipped with a nozzle. Its exit is positioned 8.5D downstream of the tip of the model nose. The nozzle body is threaded into the holder so that it can be easily replaced. Two different nozzles were designed and produced, both having a throat diameter of 4 mm, designed for Mach 1 and Mach 4 exit speeds.

A flow-through balance is integrated into the support sting, with the model mounted onto the balance via a connecting pin. This configuration prevents physical contact between the model and the support sting. Consequently, all mechanical loads acting on the model – including external aerodynamic forces and internal jet propulsion forces – are transmitted exclusively to the balance.

2.3. High-Pressure Supply System

Compressed air for powered model testing is prepared in an adjacent facility near the VTI T-38 wind tunnel. The plant utilizes two Bauer compressor units operating independently or in parallel at a maximum operating pressure of 270 bar. Initial moisture separation occurs within the compressors, followed by final drying via two aluminosilicate desiccant dryers. The conditioned air is stored at a maximum pressure of 200 bar in 14 reservoirs, providing a total volume of 9.8 m^3 (0.7 m^3 each).

The air preparation plant is linked to the VTI T-38 wind tunnel hall via a dedicated pipeline, consisting of metallic tubes with an inner diameter (ID) of 63.5 mm. Supply pressure entering the pipeline is modulated up to a maximum pressure of 150 bar using control valves located on a control panel positioned adjacent to the reservoirs.

A valve panel positioned near the wind tunnel test section enables the air flow to the model. The pipeline terminal is connected to this panel instrumentation via a 32 mm thermoplastic hose with a maximum operating pressure of 310 bar. The panel houses a computer-controlled solenoid valve (550 bar, 13 mm ID) and two manual isolation valves (350 bar, 25 mm ID) arranged in series and parallel. Upstream pressure is monitored by a 160-bar manometer. Downstream, a 25.8 mm ID thermoplastic hose (280 bar rating) routes the air to the model support mechanism, using a 6 m length for subsonic/supersonic testing or a 10 m length for transonic configurations.

Air is routed through the model support strut via a thermoplastic hose (16.3 mm ID, 140 bar rating), which connects to the upstream hose through a quick-disconnect threaded coupling (300 bar rating) for rapid assembly. The upper end of this hose exits the strut and terminates at a flexible rubber pneumatic hose (8 mm ID, 300 bar). The hose spans the junction between the fixed and moving parts of the model support. Its forward end connects to a pneumatic tube (10 mm ID, 200 bar rating) routed through an opening in the conical section of the support sting, establishing the compressed air flow path to the model.

Compressed air is delivered into the model interior via a pneumatic tube passing through the central bore of the support sting. The forward end of the tube connects to a flow-through balance. An internal axial channel (10 mm ID) along the balance body routes the airflow through a downstream tube (10 mm ID) into the settling chamber (44 mm ID, 143 mm length). The flow is then accelerated and discharged to the atmosphere through a nozzle.

2.4. Measurement Techniques and Equipment

For the jet supply system calibration, two measurement techniques were employed: jet reaction force measurement and jet supply flow measurements.

Force measurement was performed utilizing the VTI 1.5” 385 five-component flow-through internal strain-gauge balance. The normal force component (N) features a full-scale range of 2000 N with an accuracy of 0.144% of full scale; notably, only this component was relevant for the present measurement. Furthermore, calibration procedures

verified the balance accuracy under loads combined with internal airflow at a maximum pressure of 60 bar.

The jet supply flow is characterized by total pressure (p_0), static pressure (p), and total temperature (T_0). The mass flow rate is subsequently derived from these properties via standard isentropic relations.

Total and static pressures are acquired at the same axial station immediately upstream of the settling chamber using a Pitot probe and a wall pressure tap, respectively (Figure 2). These sensors are pneumatically connected to model-support-mounted miniature Entran pressure transducers, each featuring a measurement range of 70 bar and an accuracy of 0.05% of full scale.

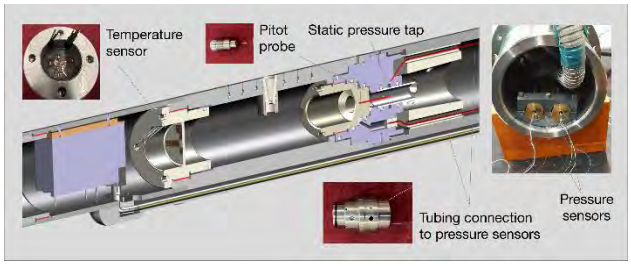


Figure 2 Jet supply flow measurements

The total temperature is measured in the settling chamber via a custom VTI half-bridge Wheatstone sensor. The circuit incorporates two active, high-pressure-rated ETG-50AE gauges (50 Ω ± 0.3%) and two precision resistors (47 Ω ± 0.1%), providing a 300 K measurement range with an accuracy of 0.2% of full scale. Dual-gauge flow exposure provides excellent common-mode rejection of strain and lead-resistance artifacts. Furthermore, the opposed-arm configuration linearizes bridge output and minimizes self-heating errors during constant-voltage operation.

2.5. Data Acquisition and Synchronization

The Pacific Instruments Series 6000 (PI6000) was utilized for high-bandwidth analog signal acquisition. All analog channels from the balance and jet supply flow sensors were conditioned and digitized by PI6000 Model 6052 input modules, each equipped with independent 16-bit A/D converters. A global sampling frequency of 400 Hz was applied across all channels. To suppress aliasing, each channel was filtered using an integrated four-pole low-pass Butterworth filter with a 30 Hz cutoff frequency.

System synchronization was managed via a PI6000 Model E6094A IRIG time-code reader, which decoded the IRIG-B signal and distributed the master sample clock, restricting intra-chassis analog skew to below 50 ns.

3. DATA PROCESSING

Evaluating the jet performance requires an accurate determination of the pneumatic state at the nozzle inlet. Compressed air routes through the internal balance into a settling chamber upstream of the nozzle. Because of frictional resistance along the internal conduits and expansion losses in the settling chamber, the pressure reaching the nozzle inlet drops substantially below the supply source pressure. To account for these losses, a mathematical methodology is developed below to calculate both the mass flow rate and the true stagnation (jet) pressure at the nozzle entry.

The mass flow rate ($\dot{m}_{in}$) delivered to the settling chamber is determined through the isentropic relations given in Eqn. (1), which utilize the jet supply flow parameters acquired just upstream of the chamber:

$$\dot{m}_{in} = \sqrt{\frac{2\kappa}{R(\kappa-1)}} A_{in} \frac{p_0}{\sqrt{T_0}} \left(\frac{p}{p_0}\right)^{\frac{1}{\kappa}} \sqrt{1-\left(\frac{p}{p_0}\right)^{\frac{\kappa-1}{\kappa}}} \tag{1}$$

where A_{in} is the inlet pipe flow area, κ is the specific heat ratio for air, and R is the specific gas constant for air. Recording these specific parameters allows for a complete thermodynamic characterization of the internal flow prior to its entry into the settling chamber.

Upon initiating the jet supple flow, the internal chamber dynamics undergo a transient phase. Steady-state behavior is achieved once the mass flow entering the chamber balances the mass flow discharging through the nozzle throat ($\dot{m}_{jet}$). Assuming choked conditions at the nozzle, this discharge mass flow becomes a function exclusively of the throat area (A_{th}) and the inlet properties (p_{jet}, T_0), as formulated in Eqn. (2):

$$\dot{m}_{jet} = \sqrt{\frac{\kappa}{R}} \left(\frac{2}{\kappa+1}\right)^{\frac{\kappa+1}{2(\kappa-1)}} A_{th} \frac{p_{jet}}{\sqrt{T_0}} \tag{2}$$

Under steady-state conditions, the settling chamber pressure (p_{sc}) represents an ideal jet pressure, derived by equating the mass flow rates from Eqns. (1) and (2). The actual stagnation pressure at the nozzle inlet (p_{jet}) is subsequently related to this ideal value by subtracting the internal pressure losses (Δp_{sc}), as expressed in Eqn. (3):

$$p_{jet} = p_{sc} - \Delta p_{sc} \tag{3}$$

Although the ideal jet pressure (p_{sc}) serves as a theoretical baseline, the true stagnation pressure (p_{jet}) at the nozzle inlet must incorporate irreversible losses within the settling chamber. To achieve an accurate determination of p_{jet}, the analytical model is coupled with empirical data by correlating the measured jet reaction force (F_{jet}) with the thrust equation, Eqn. (4):

$$F_{jet} = \dot{m}_{jet} v_e + (p_e - p_{inf}) A_e \tag{4}$$

Within this framework, p_{jet} serves as the independent variable required to close the system of equations. Given that the nozzle operates in a choked state, its exit Mach number (M_e) is first computed numerically via the isentropic area-Mach relation, Eqn. (5):

$$\frac{A_e}{A_{th}} = \left(\frac{\kappa+1}{2}\right)^{-\frac{\kappa+1}{2(\kappa-1)}} \cdot \frac{\left(1+\frac{\kappa-1}{2}{M_e}^2\right)^{\frac{\kappa+1}{2(\kappa-1)}}}{M_e} \tag{5}$$

The nozzle exit temperature (T_e) is then calculated using the resolved M_e and the measured settling chamber total temperature (T_0) via the isentropic relation, Eqn. (6):

$$T_e = T_0 \left(1 + \frac{\kappa-1}{2} M_e^2\right)^{-1} \tag{6}$$

The exit static temperature (T_e) is subsequently utilized to determine the flow velocity at the nozzle exit (v_e) via the following relation, Eqn. (7):

$$v_e = M_e\sqrt{\kappa R T_e} \tag{7}$$

By substituting these thermodynamic dependencies and the mass flow rate relation from Eqn. (2) into the thrust equation, Eqn. (4), p_{jet} is solved numerically to match the experimentally measured reaction force (F_{jet}). This approach ensures that all internal chamber losses are implicitly captured within the resolved p_{jet} value.

The nozzle exit static pressure (p_e) is calculated assuming an isentropic expansion from the jet pressure (p_{jet}) to the exit Mach number (M_e), Eqn. (8):

$$p_e = p_{jet}(1 + \frac{\kappa-1}{2}M_e^2)^{-\frac{\kappa}{\kappa-1}} \tag{8}$$

For calibration tests conducted without wind tunnel flow, the ambient pressure term (p_{inf}) in Eqn. (4) is taken as the local atmospheric pressure.

4. RESULTS

To isolate the characteristics of the jet supply circuit from external aerodynamic influences, calibration tests were conducted with the model mounted in the VTI T-38 test section under no-wind-tunnel-flow conditions (Figure 3).

Figure 3 Powered model mounted in the VTI T-38 wind tunnel

This approach ensures that the recorded pressure and temperature fluctuations are solely a function of the internal flow physics and valve actuation. Consequently, this methodology serves to calibrate the jet supply system, enabling a comprehensive evaluation of both its transient and steady-state performance characteristics.

4.1. Transient Performance

The previously established analytical framework assumes a steady-state equilibrium between the supply circuit and the nozzle exit. However, in practical wind tunnel testing, the time required to attain this equilibrium represents a critical operational constraint. This section evaluates the temporal behavior of the jet supply system, focusing specifically on the pneumatic latency and the filling dynamics of the settling chamber.

The test sequence was executed across nine discrete pressure steps, increasing the supply pressure (p_s) from 20 bar to 60 bar in 5 bar increments. For each step, data acquisition was triggered simultaneously with the opening of the computer-controlled solenoid valve. The measured quantities were continuously recorded for a 10-second duration, ensuring a sufficient window to capture the entire transition from the initial rest state to a stabilized steady-state plateau. All data channels were acquired at a sampling frequency of 400 Hz. This entire testing sequence was first performed for the Mach 1 nozzle configuration and subsequently repeated for the Mach 4 nozzle.

Figure 4 illustrates the transient mass flow rate ($\dot{m}_{in}$) response for the Mach 1 nozzle configuration across the evaluated range of supply pressures (p_s). To suppress experimental noise while preserving the transient peak characteristics, the raw time-series data were conditioned using a second-order Savitzky-Golay filter with a 50-point moving window (corresponding to 0.125 s).

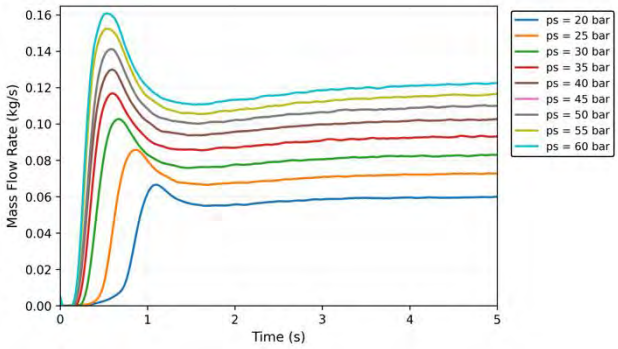


Figure 4 Transient response of the jet supply circuit

Comparative testing revealed no significant discrepancy in transient response between the Mach 1 and Mach 4 nozzle configurations. This indicates that pneumatic latency and filling dynamics are dominated by the upstream supply circuit and chamber volume rather than nozzle geometry.

The time-dependent mass flow rate ($\dot{m}_{in}$) exhibits several distinct operational phases. Following the valve trigger, a distinct dead time occurs before initial flow detection. This latency is inversely proportional to the supply pressure – decreasing from 0.5 s at 20 bar to 0.2 s at 60 bar – and is attributed to the pressure wave propagation time through the internal piping. Post-latency, the flow exhibits a steep gradient ($d\dot{m}_{in}/dt$) leading to a significant transient overshoot. At 60 bar, the mass flow peaks at approximately 0.16 kg/s, which is nearly 33% higher than its eventual steady-state value.

This overshoot stems directly from the volumetric filling of the settling chamber. Initially, the maximum pressure differential between the supply line and the empty chamber drives a high inflow, while the nozzle exit flow is restricted by the low chamber pressure. As the chamber fills, the rising internal pressure (p_{sc}) simultaneously reduces the inflow rate and increases the nozzle discharge rate. Equilibrium is established roughly 4.0 s after the initial rise, when the filling and exhaust mass flows balance.

Consequently, this dynamics dictate the wind tunnel testing protocol. To ensure true steady-state conditions, a minimum dwell time of 4.0 s is required post-activation before recording valid aerodynamic data. This delay ensures the pneumatic system has fully transitioned into the steady-state regime assumed by the analytical model.

4.2. Steady-State Performance

This section characterizes the steady-state pneumatic efficiency of the jet supply circuit. By coupling the analytical framework from Section 3 with empirical force and pressure data, internal losses between the upstream supply lines and the settling chamber are quantified for both the Mach 1 and Mach 4 nozzle configurations.

The tests were conducted across nine discrete supply pressure steps (p_s= 20 bar to 60 bar in 5 bar increments). Steady-state data acquisition simultaneously captured the total duct pressure (p_0), static duct pressure (p), total temperature (T_0), and jet reaction force (F_{jet}). To exclude the transient filling dynamics, data logging was initiated only after a 4-seconds post-activation dwell time. For each pressure step, parameters were sampled at 400 Hz for 10 s. This approach yielded a statistically robust dataset that minimized high-frequency pneumatic noise and ensured a precise determination of the mean properties.

Prior to testing, system integrity was verified via a pressure-decay leak check. The isolated circuit was pressurized to 60 bar and monitored for 10 minutes using the internal transducers (p_0, p); no detectable pressure drop occurred. This verified zero mass leakage into the model interior, preserving force measurement accuracy.

The Mach 1 and Mach 4 nozzle comparative performance at tested supply pressures are summarized in Figure 5. The total pressure drop across the system (Δp_{tot}) is decomposed into two primary components: losses within the piping upstream of the model (Δp_{up}) and losses generated inside the settling chamber itself (Δp_{sc}). The latter is resolved by comparing the ideal jet pressure (Δp_{sc}) with the actual jet stagnation pressure (Δp_{jet}) according to Eqn. (3) of the analytical model.

Although the upstream piping constitutes the largest portion of the total pressure drop, the settling chamber contribution remains significant. Specifically, for the Mach 1 nozzle, the settling chamber's share of total losses increases from 21.0% at p_s= 20 bar to 31.7% at p_s= 60 bar. This trend indicates that internal chamber inefficiencies driven by intensified turbulence and irreversible expansion become increasingly dominant at higher mass flow rates.

The jet stagnation pressure p_{jet} as a function of supply pressure p_s for both nozzle configurations is presented in Figure 6, together with the ideal settling chamber pressure p_{sc} calculated from mass-flow continuity. At the lowest supply pressure (p_s = 20 bar), both nozzles deliver nearly identical jet stagnation pressures (p_{jet} ≈ 14.8 bar). As supply pressure increases, however, a clear divergence emerges. The Mach 4 nozzle consistently achieves higher values of p_{jet} than the Mach 1 nozzle, with the difference becoming progressively larger above 40 bar. At p_s = 60 bar, the Mach 4 nozzle reaches p_{jet} ≈ 44.7 bar compared to 42.6 bar for the Mach 1 nozzle.

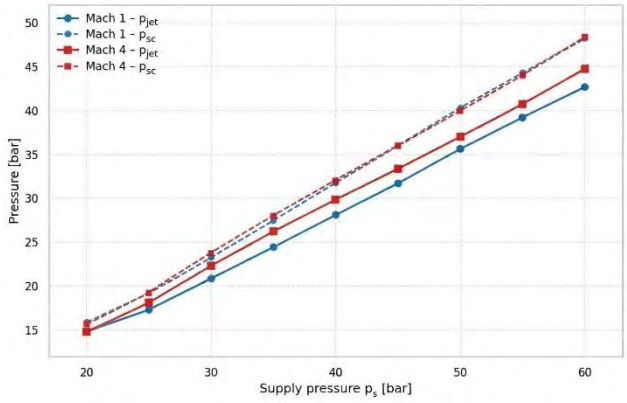


Figure 6 Comparative pressure recovery analysis

This performance advantage is directly attributable to differences in internal flow behavior within the settling chamber. As shown in Figure 5, the absolute internal pressure losses ($\Delta p_{up} + \Delta p_{sc}$) are lower for the Mach 4 nozzle at higher supply pressures, despite nearly identical upstream losses in both configurations. The contribution of settling chamber losses (orange segments) is visibly smaller for the Mach 4 geometry across the entire range, and the gap widens with increasing mass flow rate. Consequently, the deviation between the measured jet stagnation pressure p_{jet} and the ideal settling chamber pressure p_{sc} remains smaller for the supersonic nozzle, indicating reduced irreversible losses associated with flow expansion and turbulence inside the chamber. In addition, this deviation demonstrates that internal flow losses are successfully captured by the force-based method.

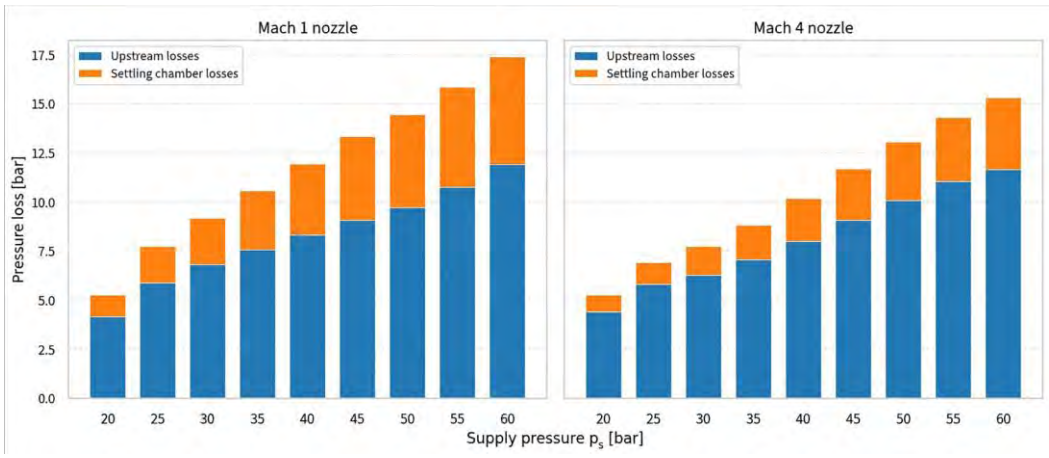


Figure 5 Pressure loss contribution analysis

Figure 7 presents the directly measured jet reaction force F_{jet} as a function of the jet stagnation pressure p_{jet} determined from the thrust equation for both nozzle configurations under atmospheric (no-wind-tunnel-flow) conditions. A clear linear relationship is observed for both the Mach 1 and Mach 4 nozzles, consistent with the theoretical thrust equation for choked nozzles. The Mach 4 nozzle exhibits a steeper slope, producing higher reaction force at the same jet stagnation pressure due to its higher exhaust velocity resulting from supersonic expansion in the divergent section. The data points remain well within the uncertainty bands across the entire tested range. The two curves intersect at low jet stagnation pressures; below this intersection the Mach 1 nozzle yields slightly higher force, while above approximately 20 bar the trend reverses and the Mach 4 nozzle maintains progressively higher reaction force with increasing p_{jet}.

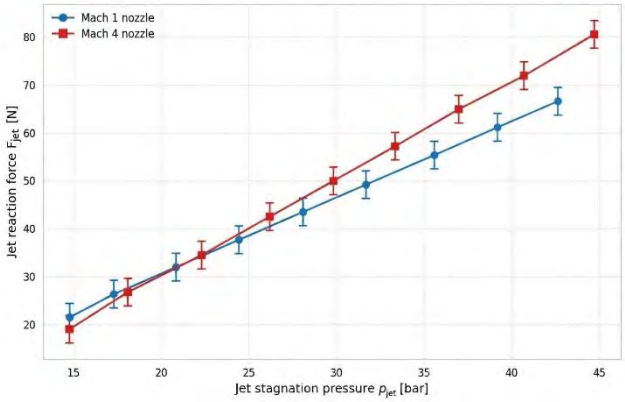


Figure 7 Jet reaction force under atmosphere conditions

In the wind tunnel, the nozzle exhausts into a local test section static pressure that can differ significantly from the atmospheric pressure present during the static calibration. Consequently, the jet thrust (F_{jet}) obtained here must be adjusted analytically using the classical thrust equation. In the pressure thrust component $(p_e - p_{inf})A_e$, the atmospheric reference pressure is replaced with the actual local test section pressure, allowing the true jet thrust force to be isolated in the subsequent wind tunnel tests.

5. CONCLUSIONS

The high-pressure jet supply system was calibrated under no-main-flow conditions using direct jet reaction force measurements. A dwell time of 4 s was required to reach steady-state conditions in the internal settling chamber. Upstream piping losses represented the dominant contribution to the total internal pressure drop for both nozzle configurations. Settling chamber losses were lower for the Mach 4 nozzle and increased with supply pressure for both nozzles. The Mach 4 nozzle delivered higher jet stagnation pressure and jet reaction force than the Mach 1 nozzle at supply pressures above 25 bar. A linear relationship between jet reaction force and jet pressure was obtained for both nozzles, consistent with the theoretical thrust equation. The force-based method enabled determination of actual jet stagnation pressure, providing validated baseline data for powered wind tunnel testing.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No 451-03-34/2026-03/200325).

REFERENCES

[1] A.A. Dyakonov, M. Schoenenberger, W.I. Scallion, J.W. Van Norman, L.A. Novak, C.Y. Tang, Aerodynamic Interference Due to MSL Reaction Control System, in: 41st AIAA Thermophysics Conference, San Antonio, 2009.

[2] B. Stahl, H. Emunds, A. Gülhan, Experimental investigation of hot and cold side jet interaction with a supersonic cross-flow, Aerosp. Sci. Technol. 13 (2009) 488–496.

[3] A. Kovar, E. Schülein, Effect of Side Jets in a Supersonic Flow Measured and Calculated on a Flat Plate and a Generic Missile Configuration, Innovative Missile Systems Paper 37 (2006) 1–16.

[4] C.T. Johansen, L. Novak, B.F. Bathel, S.W. Ashcraft, P.M. Danehy, Mars Science Laboratory Reaction Control System Jet Computations with Visualization and Velocimetry, J. Spacecr. Rockets 50 (2013) 1183–1195.

[5] J. Melton, K. James, J.D. Flamm, K. Long, Estimating Flow-Through Balance Momentum Tares with CFD (Invited), in: 54th AIAA Aerospace Sciences Meeting, American Institute of Aeronautics and Astronautics, 2016.

[6] D.E. Burns, P.A. Parker, C.M. Cagle, K.G. Toro, B. Kleb, V.M. Hawke, New Flow-Through Wind-Tunnel Balance for Retropropulsion Testing, J. Spacecr. Rockets 59 (2022) 1926–1933.

[7] B. Ilić, M. Samardžić, G. Ocokoljić, Integrated multi-technique experimental setup for wind tunnel investigation of side jet interaction with crossflow, Aerosp. Sci. Technol. 177 (2026) 112705.

[8] B. Ilić, G. Ocokoljić, M. Samardžić, J. Isaković, Upgrade of the VTI T-38 Blowdown Trisonic Test Facility: Wind Tunnel Control System, in: ICAS Proceedings - 34th Congress of the International Council of the Aeronautical Sciences, Florence, 2024, ICAS2024_0456.

[9] B. Ilić, M. Miloš, J. Isakovic, Cascade nonlinear feedforward-feedback control of stagnation pressure in a supersonic blowdown wind tunnel, Measurement 95 (2017) 424–438.

[10] B. Ilić, M. Milosavljević, J. Isaković, M. Miloš, Stagnation Pressure Transient Control in a High-Pressure Supersonic Blowdown Wind Tunnel Test Facility, 32nd DAS, Stary Smokovec, 2015: pp. 82–83.

[11] B. Ilić, M. Miloš, M. Milosavljević, J. Isaković, Model-Based Stagnation Pressure Control in a Supersonic Wind Tunnel, FME Transactions 44 (2016) 1–9.

[12] B. Ilić, M. Milosavljević, S. Krstić, G. Ocokoljić, D. Damljanović, Control System Design and Integration of a Blowdown Wind Tunnel Model Support Pitch Mechanism, OTEH 2022 - Proceedings, Belgrade, 2022: pp. 210–216.

[13] D. Vuković, B. Ilić, M. Milosavljević, Upgrade of the Roll-Angle-Setting Mechanism on the Model Support System in the T-38 Wind Tunnel in VTI, OTEH 2011 - Proceedings, 2011: pp. 23–28.

[14] B. Ilić, M. Milosavljević, FPGA-based Embedded System for Wind Tunnel Variable-Geometry Nozzle Positioning, Scientific Technical Review 69 (2019) 3–9.

[15] G.M. Elfstrom, B. Medved, W.J. Rainbird, Wave cancellation properties of a splitter-plate porous wall configuration, J. Aircr. 26 (1989) 920–924.

[16] M. Samardžić, N. Vujić, B. Ilić, G. Ocokoljić, Experimental Results from the Missile Model Test Campaign in the VTI T-38 Wind Tunnel, OTEH 2024 - Proceedings, pp. 43–49.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARISON OF MISSILE AERODYNAMIC DATA OBTAINED IN WIND TUNNEL TESTS USING TWO INTERNAL STRAIN GAUGE BALANCES

MARIJA SAMARDŽIĆ
Military Technical Institute, Belgrade, marija.samardzic.vti@gmail.com

BILJANA ILIĆ
Military Technical Institute, Belgrade, biljana.ilic@mod.gov.rs

ISIDORA STANKOVIĆ
Military Technical Institute, Belgrade,stank.151dora@gmail.com

BOŽIDAR SIMOVIĆ
Military Technical Institute, Belgrade,simovic.bozidar@gmail.com

OGNJEN BABIĆ
Military Technical Institute, Belgrade, ognjen813@gmail.com

GORAN OCOKOLJIĆ
Military Technical Institute, Belgrade,goran.ocokoljic@vti.vs.rs

Abstract: *This paper presents experimental data obtained in testing of the generic missile model in the VTI T-38 blowdown wind tunnel. Two internal strain gauge balances were used to measure aerodynamic load. Force-type six-component balance with two measurement sections, placed in both the forward and the aft section of the balance, was used in the initial part of the tests. In the subsequent tests, aerodynamic load was measured using monoblock moment - type five-component balance with three measuring section shaped as a five-bar cage. The five-component balance has cylindrical cavity that allows flow of high-pressure gas through its load-bearing parts. The tests were conducted in the subsonic, transonic and supersonic conditions. The aerodynamic load values measured using these two balances are compared for the angle of attack of the missile model in the range from -12° to +12°.*

Keywords: *wind tunnel, strain gauge balance, missile model, aerodynamic load*

1. INTRODUSCTION

The paper presents the results of testing the generic missile model at VTI T-38 blowdown wind tunnel [1-3]. The main aim of the tests was investigation of the side jet interactions with crossflow. Tests were done in two phases: in the initial phase of the testing, i.e. in the first phase of the testing, measurements of the aerodynamic forces and moments were conducted on the generic missile model without side jet system and with six-component internal wind tunnel balance [4]. In the second phase of testing, measurements of aerodynamic forces and moments were conducted with side jet system and with five-component flow-through balance. Nominal load ranges of the flow-through balance, used in the tests with side jet system, is significant lower compared to the nominal load ranges of the six-component balance. The experimental results obtained in tests with six-component balance were used to determine the optimum position of the five-component balance within the model and maximum Mach number in supersonic testing of the model with side jet system.

During the second phase of the testing a certain number of tests were done without side jet system operating. Aerodynamic loads measured in these tests represent the aerodynamic loads without the influence of the jet's reactive force. Also, these tests without the side jet system operating were used as control measurements for comparing the aerodynamic loads measured using the six-component balance and aerodynamic loads measured using flow-through balance. The paper presents comparative values of the aerodynamic loads measured using both wind tunnel balances without side jet system operating in the subsonic, transonic and supersonic speed ranges.

2. GENERIC MISSILE MODEL

Generic missile model has a cylindrical body, ogive nose and wing section with for wraparound wings. The model is characterized with large slenderness ratio. Ratio of the model length to the model base diameter is L/D=21.6.

Generic missile model, in the initial phase of the testing with six-component balance, in the VTI T-38 wind tunnel is shown in Figure 1.

Corresponding author: M. Samardžić;
E-mail:marija.samardzic.vti@gmail.com
DOI: 10.5937/OTEH26006S

Figure 1. Generic missile model in the VTI T-38 wind tunnel test section [4]

For the tests with the side jet system, it was necessary to make some modifications on the front part of the model. Modification of the model baseline geometry enabled high pressure air supply from through flow-through balance to the settling chamber installed inside the model. On the upper surface of the settling chamber two interchangeable nozzles, a sonic (Mach 1) nozzle and supersonic (Mach 4) nozzle, can be installed. Generic missile model on the flow-through five-component balance are shown in Figure 2.

Figure 2. Generic missile model on the flow-through five component balance

Center of the six-component balance was located at 13.3*D* downstream of the model nose tip, while center of the flow-through balance was located at 13.7*D* downstream of the model nose tip. The reference point for force/moment calculation is located at 10.9*D* downstream of the model nose tip.

3. DESCRIPTION OF THE WIND TUNNEL BALANCES

3.1 Six-component wind tunnel balance

Six-component internal wind tunnel balance is shown in Figure 3. This is assembled balance with 50.8 mm-diameter (2 inch), marked as 2″Mk XVIII. This balance has construction and arrangement of measuring elements that are standard for balances produced by Able Corporation, USA. The 2″Mk XVIII balance is "force" balance, measuring five components of force and one moment component. From these six balance components six components of the aerodynamic load can be determined: the three total forces and three total moments. Two normal force elements (aft normal force element and forward normal force element) are used for measurement of the total normal force and pitching moment. Also, two side force elements (aft side force element and forward side force element) are used for measurement of the total side force and yawing moment. Rolling moment and axial force are directly measured using a dual roll element, i.e. dual axial element. The measuring bridges of all measuring elements are formed from Micro-Measurements strain gauges with a resistance of 350Ω. Rolling moment gage section is torque tube type, while all other gauge sections are bending beam type.

The nominal balance components load range is given in Table 1. Load range of the balance components relative to the balance center is given in Table 2 in the form of the "load rhombs", i.e. as maximum side/normal force at zero yawing/pitching moment and as maximum yawing/pitching moment at zero side/normal force.

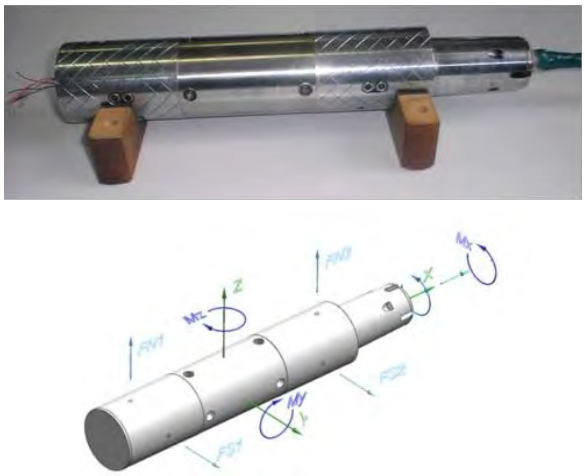


Figure 3. Six-component 2″ Mk XVIII balance

Table 1. Six-component balance load ranges

2″ Mk XVIII balance components load ranges	
Axial force, *X*	2699 N
Forward normal force, *FN1*	8896 N
Aft normal force, *FN2*	8896 N
Forward side force, *FS1*	8896 N
Aft side force, *FS2*	8896 N
Rolling moment, *Mx*	340 Nm

Table 2. Six-component balance load ranges relative to the balance center

2″ Mk XVIII balance components load ranges relative to the balance center	
Axial force, *X*	2699 N
Side force, *Y*	17792 Nm at *Mz=0* Nm, 0 N at *Mz*=1356 Nm
Normal force, *Z*	17792 Nm at *My=0* Nm, 0 N at *My*=1638 Nm
Rolling moment, *Mx*	340 Nm
Pitching moment, *My*	1638 Nm at *Z=0* N, 0 Nm at *Z*=17792 Nm
Yawing moment, *Mz*	1356 Nm at *Y=0* N, 0 Nm at *Y*=17792 Nm

The balance accuracy obtained in a checkout after the

calculation of the calibration matrix is given in Table 3. Accuracy is interpreted as equal to two standard deviations of the back-calculated error for all data points in the calibration process (corresponding to 95% measurements certainty).

Table 3. Six-component balance accuracy

2″ Mk XVIII balance accuracy (balance components)						
	FN1 [N]	*FN2* [N]	*FS2* [N]	*FS2* [N]	*Fx* [N]	*Mx,* [Nm]
Err, [N]	13.412	21.522	-11.704	-14.170	3.467	-0.406
P.Err, [%FS]	0.151	0.242	-0.132	-0.159	0.193	-0.239
Std.d, [%FS]	±0.039	±0.049	±0.041	±0.045	±0.053	±0.051
Acc, [%FS]	±0.078	±0.098	±0.082	±0.090	±0.106	±0.102
2″ Mk XVIII balance accuracy (relative to the balance center)						
	X [N]	*Y* [N]	*Z* [N]	*Mx* [Nm]	*My* [Nm]	*Mz* [Nm]
Err, [N]	3.467	23.214	34.303	-0.406	0.887	1.114
P.Err, [%FS]	0.193	0.261	0.385	-0.239	0.108	0.164
Std.d, [%FS]	±0.053	±0.075	±0.083	±0.051	±0.029	±0.041
Acc, [%FS]	±0.106	±0.150	±0.166	±0.102	±0.058	±0.082

In Table 3 Err is maximum difference between applied and measured load in the balance calibration, P.Err is maximum difference between applied and measured load calculated in relation to the component full scale (FS) load range, Std.d is standard deviation of the errors, Acc is balance component accuracy.

3.2 Five-component flow-through wind tunnel balance

Five-component flow-through wind tunnel balance is shown in Figure 4. The VTI 1.5” 385 balance is “direct-read” balance with 38 mm-diameter (1.5 inch) and three measuring sections shaped as a five-bar cage. All measuring bridges are formed from Micro-Measurements strain gauges with a resistance of 350Ω. The normal force and pitching moment, as well as side force and yawing moment are measured using four measuring bridges formed from strain gauges located on the forward and backward measuring sections. Output signals from these four measuring bridges are proportional to the bending stresses in the measuring sections. Half bridges on the forward and backward measuring section are directly wired to the force bridges for measurements of normal and side force, while the other set of half bridges are directly wired to the moment bridges for measurements of pitching and yawing moment. Rolling moment measuring bridge is located on the central measuring section [5]. Due to action of the rolling moment, the four external bars are simultaneously twisted and bended, while the central bar is loaded only in rolling. By the balance design, it is ensured that the bending stress in the external bars is more significant than the twisting stress. The part of rolling moment taken by the central bar is a small portion of the total rolling moment. The appropriate geometry of the central measuring section enables indirect measurement of the rolling moment by bending stress in the external bars. One strain gauge that detects bending stress is located on each external bar.

High-pressure gas for a side jet flows through a cylindrical cavity in the balance, Figure 4. The cylindrical cavity is 10 mm diameter and it is threaded at both ends to connect the high-pressure gas supply tube.

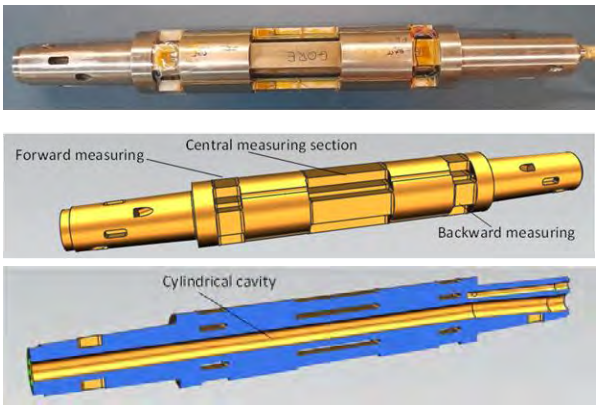


Figure 4. Five-component flow-trough wind tunnel balance

The nominal balance load range relative to the balance center is given in Table 4.

Table 4 Five-component balance load ranges

VTI 1.5” 385 balance components load range relative to the balnce center	
Side force, *Y*	3500 N
Normal force, *Z*	3500 N
Rolling moment, *Mx*	150 Nm
Pitching moment, *My*	350 Nm
Yawing moment, *Mz*	350 Nm

The VTI 1.5” 385 balance accuracy obtained in a checkout after the calculation of the calibration matrix is given in Table 5.

Table 5 Five-component balance accuracy

VTI 1.5” 385 balance accuracy					
	Y [N]	*Z* [N]	*Mx* [Nm]	*My* [Nm]	*Mz* [Nm]
Err, [N]	-4.258	-5.122	-0.389	-0.438	-0.559
P.Err, [%FS]	-0.194	-0.233	-0.324	0.199	-0.254
Std.d, [%FS]	±0.073	±0.072	±0.111	±0.058	±0.069
Acc, [%FS]	±0.156	±0.144	±0.222	±0.116	±0.138

4. EXPERIMENTAL RESULTS AND DISCUSSION

Aerodynamic loads, measured in the subsonic, transonic and supersonic speed range, with both wind tunnel balances are presented. In the subsonic speed range stagnation pressure in the wind tunnel tests section was

2.0 bar, while in transonic and supersonic speed range stagnation pressure was 2.3 bar. The results refer to Mach number 0.3, 0.5, 0.9, 1, 1.2, 1.5 and 1.6.

Nondimensional aerodynamic coefficients in the non-rotated axes system are calculated from the aerodynamic forces and moments according the following equations:

$$C_N = \frac{Z}{q \cdot S_{ref}} \quad C_M = \frac{M_y}{q \cdot S_{ref} \cdot l_{ref}}$$

where C_N is normal force coefficient, C_M is pitching moment coefficient, S_{ref} is model reference area, l_{ref} is model reference length, q is dynamic pressure in the wind tunnel test section.

The model angle of attack was in the interval -12° to +12°. Pitching moment coefficient is calculated for the reference point located at 10.9 *D* downstream of the model nose tip.

Figures 5-11 show measured values of the normal force and pitching moment coefficients in all three speed ranges.

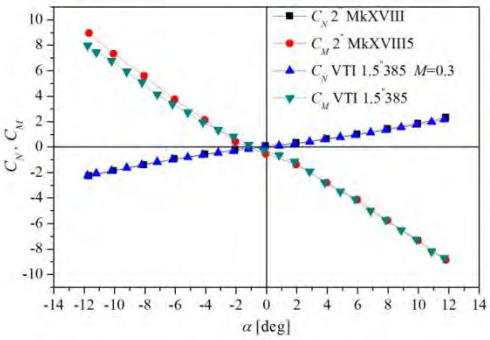


Figure 5. Normal force and pitching moment coefficients, $M = 0.3$

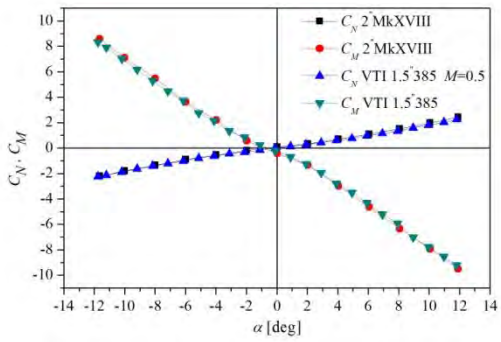


Figure 6. Normal force and pitching moment coefficients, $M = 0.5$

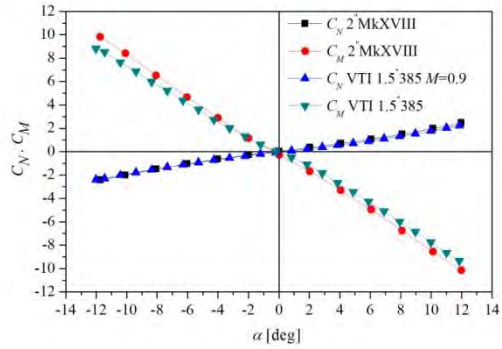


Figure 7. Normal force and pitching moment coefficients, $M = 0.9$

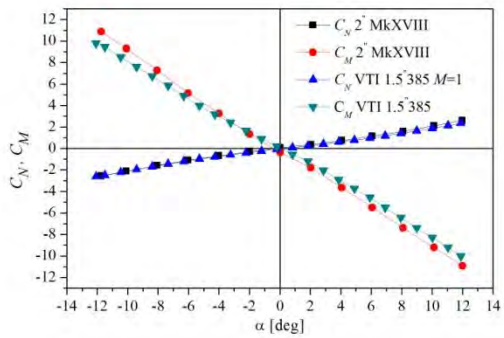


Figure 8. Normal force and pitching moment coefficients, $M = 1$

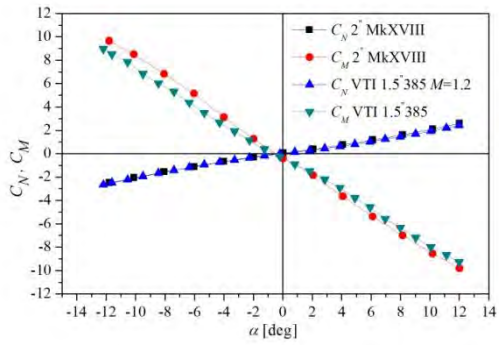


Figure 9. Normal force and pitching moment coefficients, $M = 1.2$

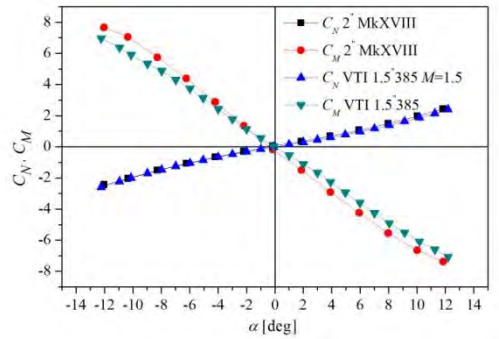


Figure 10. Normal force and pitching moment coefficients, $M = 1.5$

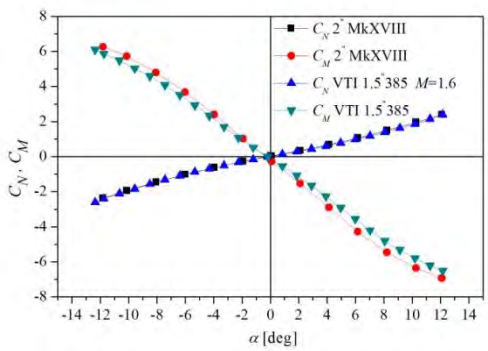


Figure 11. Normal force and pitching moment coefficients, M = 1.6

Figures 12-15 show comparative values of the normal force measured using 2″Mk XVIII and VTI 1.5” 385 wind tunnel balances.

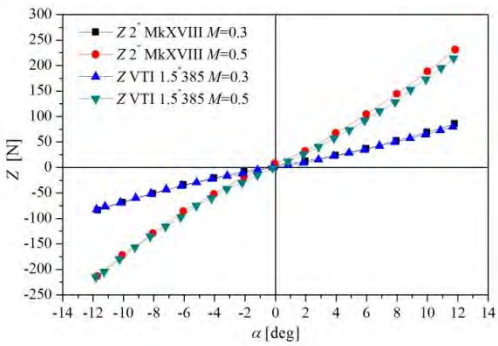


Figure 12. Normal force M = 0.3 and M = 0.5

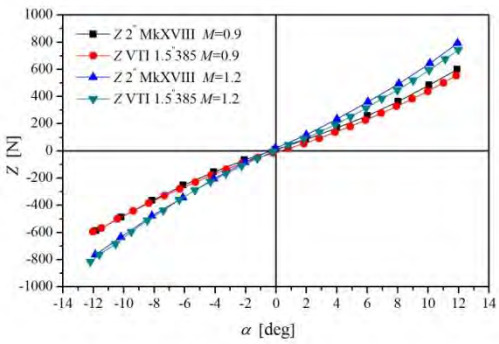


Figure 13. Normal force M = 0.9 and M = 1.2

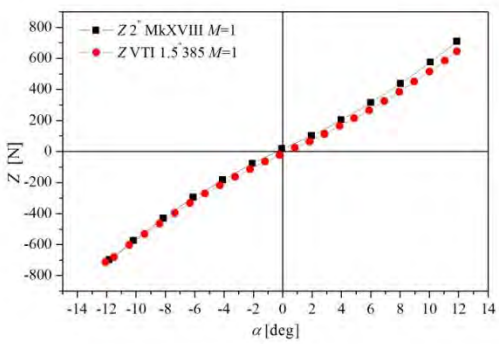


Figure 14. Normal force M = 1

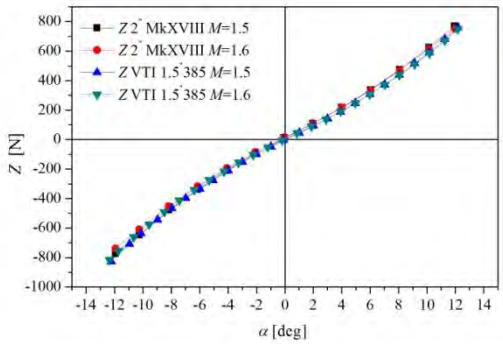


Figure 15. Normal force M = 1.5 and M = 1.6

5. CONCLUSIONS

The diagrams in Figures 5-15 show a good agreement between the normal forces measured using 2″Mk XVIII balance and VTI 1.5” 385 balances. Slightly larger difference in the normal force values can be observed in the transonic speed range. At Mach number M=1, and for the model maximum angle of attack, the difference in normal force measured values is 63.74 N. These deviations can be explained by the significantly larger 2″Mk XVIII balance normal force load range compared to the measured value.

A certain difference in the measured values of the pitching moment coefficient can be noticed at higher values of the model angle of attack. When analyzing these differences, it should be taken into account the significantly larger 2″Mk XVIII balance pitching moment load range. Also, the pitching moment measured using 2″Mk XVIII balance are calculated based on the measured forward and aft normal force components. On the other hand, using the VTI 1.5” 385 balance, the pitching moment is measured directly, i.e. it is achieved using a single measuring bridge.

Acknowledgements

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-34/2026-03/200325).

References

[1] MEDVED, B., ELFSTROM, G.: The Yugoslav 1.5 m trisonic blowdown wind tunnel, AIAA 86-0746, in: 14th Aerodynamic Testing Conference, American Testing Conference, American Institute of Aeronautics, 1986, https://doi.org/10.2514/6.1986-746.

[2] ILIĆ, B., OCOKOLJIĆ, G., SAMARDŽIĆ, M., ISAKOVIĆ, J.: Upgrade of the VTI T-38 blowdown trisonic test facility: Wind tunnel control system, ICAS Proceedings -34th Congress of the International Council of the Aeronautics Sciences, Florence, 2024: p. ICAS2024_0456 (2024).

[3] ILIĆ, B., MILOSAVLJEVIĆ, M., KRSTIĆ, S., OCOKOLJIĆ,G., DAMLJANOVIĆ, D.:Control system desigh and integration of blowdown wind tunnel model support pitch mechanism, 10th International Conference on Defensive

Technologies, OTEH 2022 Proceedings, Belgrade, 2022, pp.210-216.

[4] SAMARDŽIĆ, M., VUJIĆ, N., ILIĆ, B., OCOKOLJIĆ, G.: Experimental results from the missile model test campaign in the VTI T-38 wind tunnel, 11thInternational Conference on Defensive Technologies, OTEH 2024 Proceedings, Tara, Serbia, 2024, pp.43-49.

[5] BOUTEMDJET, A., SAMARDŽIĆ, M., ĆURČIĆ, D., RAJIĆ, Z., OCOKOLJIĆ, G.: Wind tunnel measurement of small values of rolling moment using six-component strain gauge balance, Measurement, 116, 2018, pp. 438-450, https://doi.org/10.1016/j. measurement.2017.11.043.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SIDE JET INTERACTION WITH SUBSONIC, TRANSONIC AND SUPERSONIC CROSSFLOW: SURFACE PRESSURE DISTRIBUTION

BOŽIDAR SIMOVIĆ
Military Technical Institute, Belgrade, simovic.bozidar@gmail.com

BILJANA ILIĆ
Military Technical Institute, Belgrade, biljana.ilic@icloud.com

MARIJA SAMARDŽIĆ
Military Technical Institute, Belgrade, marija.samardzic.vti@gmail.com

ISIDORA STANKOVIĆ
Military Technical Institute, Belgrade, stank151.dora@gmail.com

OGNJEN BABIĆ
Military Technical Institute, Belgrade, ognjen813@gmail.com

GORAN OCOKOLJIĆ
Military Technical Institute, Belgrade, ocokoljic.goran@gmail.com

Abstract: *This paper examines the interaction between a free stream and side jet flow. Experiments were conducted in the VTI T-38 trisonic wind tunnel. The internal side jet system was configured with either sonic or supersonic nozzles, integrated on a generic missile model. Tests were performed at the free stream Mach numbers ranging from subsonic through transonic to supersonic regimes. This interaction generated unique boundary layer profiles and oblique shock wave structures. The resulting tests yielded evaluated pressure distributions and quantified data regarding the interference between the free stream and side jet flows. Such data is essential for analyzing the side jet effects on flow separation, the location of the center of pressure, and the overall maneuverability and stability of missiles.*

Keywords: *trisonic wind tunnel; side jet; pressure distribution; flow interference and separation.*

1. INTRODUCTION

Control of missile systems designed for high agility and maneuverability is increasingly being achieved using side jet control. In contrast to control surfaces that become ineffective at low pressures, side jet control maintains its effectiveness under these conditions. The side jet's effects depend primarily on: the diameter of the side jet nozzle, jet pressure, the Mach number and static pressure of the external flow. In aerodynamic laboratories, various methods and equipment are being developed for wall pressure measurements to investigate the interference of a side jet with cross flow [1-3].

This paper presents results of the pressure distribution measurements on the generic missile model with side jets. Measurements were conducted in the VTI T-38 trisonic blowdown wind tunnel in the subsonic, transonic and supersonic speed range [4]. Tests were conducted with a sonic and supersonic side jet nozzle installed on the generic missile model and for two different values of the side jet mass flow.

2. TEST FACILITY

The VTI T-38 test facility is large scale blowdown wind tunnel. Depending on the Mach number and stagnation pressure, the facility permits run duration from 6 to 60 seconds. The 1.5 m x 1.5 m test section is equipped with a two-degree-of-freedom model support system that permits pitch angles of ±20° and roll angles of 720°. Stagnation pressure can be in range from 1.2 to 15 bar, enabling unit Reynolds numbers, R_e/m, from 5 x 10^6 m^{-1} to 110 x 10^6 m^{-1}. Mach number range in the VTI T-38 wind tunnel test section is 0.2-4.0. Subsonic and supersonic tests are performed in the solid-wall test section, where the Mach number is set by second throat and the variable-geometry nozzle, respectively. For transonic tests, a porous-wall test section is with variable porosity and an ejector-assisted blow-off system is inserted into the wind tunnel circuit for active Mach number control.

2.1 Model description

The generic missile model is developed in VTI, Figure 1. This model is designed for testing primarily in supersonic speed range. The generic missile model consists of an

Corresponding author: B. Simović;
E-mail: simovic.bozidar@gmail.com
DOI: 10.5937/OTEH26007S

ogive nose cone, cylindrical fuselage and four wrap-around fins. Its nominal body dimeter is D = 64 mm. The total length of the cylindrical fuselage, which consists four segments, is 18 D. The ogive nose cone has a length of 3.6 D. The generic missile model is characterized with large slenderness ratio. Ratio of the model length to the model base diameter is close to L/D = 22. Four movable wrap-around fins are installed on the rear fuselage segment and can be individually set to different deflection angles. For the present test campaign, the fins were positioned at 0° deflection. In the initial phase of the testing the model was tested without side jet system and with six-component internal wind tunnel balance [6]. In the measurements with side jet system, aerodynamic forces and moment were measured with five-component flow-through balance. The experimental results obtained in tests with six-component balance were used to determine the optimum position of the five-component balance within the model and maximum Mach number in supersonic testing of the model with side jet system.

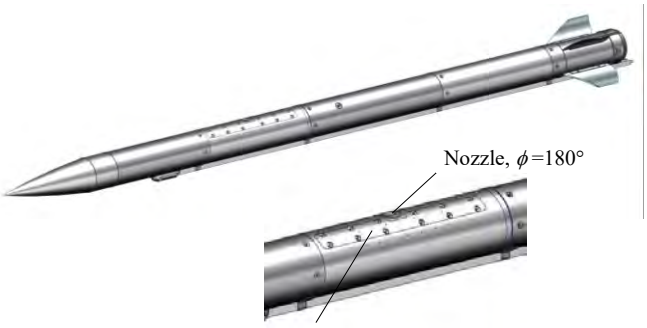


Figure 1. The generic missile model

For investigation of side jet interactions with cross-flow two interchangeable side jet nozzles were designed and manufactured: a sonic (Mach 1) and supersonic (Mach 4) nozzle, both with throat diameter of d = 4 mm. The side jets are positioned at 8.5 D downstream of the model nose and at an azimuth angle of ϕ = 180°. The side jet nozzles are supplied with high-pressure air from an external source through five-component flow-through balance and internal settling chamber that houses the nozzles [7].

2.2 Pressure distribution measurements

For measurements of the wall pressure distributions, the model has 34 static pressure taps with a diameter of 0.8 mm. The static pressure taps are arranged in five longitudinal rows around the nozzle exit, Figure 1. The pneumatic tubing from the taps are connected to a Scanivalve MPS4264 miniature digital pressure scanner module, Figure 2. The MPS4264 is a 64-channel scanner with a full-scale range of 1 bar, long-term accuracy of ±0.06 % of full scale, integrated 24-bit analog-to-digital conversion, and maximum scan rate 850 Hz. Reference pressure for all scanner channels is static pressure measured in the tunnel tests section, i.e. pressure of the external flow.

Each scanner channel directly outputs the pressure difference between local surface static pressure and static pressure of the external flow, p-p_∞ , which is used to calculate the pressure coefficient according to next equation:

$$Cp = 2\,(p-p_\infty)/(p_\infty k M^2) \qquad (1)$$

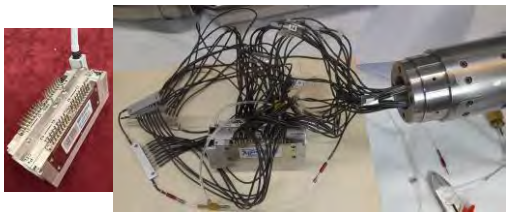


Figure 2. Miniature pressure scanner module

By varying the external source pressure, it was possible to set a predefined side jet pressure ratio p_{ojet}/p_∞. The pressure ratio of the side jet was set to a constant value before the start of cross flow in the wind tunnel test section. The model was traversed in 5° steps in the range -10°≤ α ≤ 10°, or through -12°≤ α ≤ 12° angles during more dynamic and faster sweep regime of measurement. The shock structure is visualized by schlieren technique. The schlieren system for the VTI T-38 wind tunnel is a Toepler parallel beam folded Z-type system.

3. EXPERIMENTAL RESULTS AND DISCUSION

The generic missile model with side jet system in the VTI T-38 wind tunnel test section is shown in Figure 3.

Figure 3. The generic missile model in the VTI T-38 wind tunnel

Experimental results of the side jet interaction with cross flow for Mach numbers M = 0.5, 0.9 and 1.6 are presented. The results refer on sonic nozzle (Mach 1) and the model angle of attack 0°.

The measurements were carried out under test conditions shown in Table 1.

Table 1. Test condidtions

M	M_{jet}	Re	p_o [bar]	p_∞ [bar]	p_{ojet}/p_∞
0.5	1	2.15×10^7	2.0	1.675	18.55
0.9	1	3.06×10^7	2.3	1.360	23.955
1.6	1	3.24×10^7	2.3	0.561	49.72

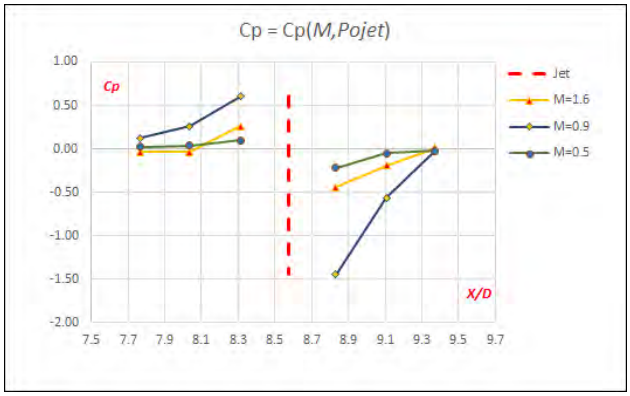


Figure 4. Pressure coefficient distribution: dependence on Mach number and side jet flow characteristics

Pressure distribution across different Mach numbers, shown in figure 4, exhibits close similarity in its overall curve character. Observation reveals that free flow loses velocity (indicated by higher *Cp* values) as it approaches the jet. This interaction between the free flow and the jet effectively acts as a barrier, reducing dynamic pressure and contributing to the overall force generated by the side jet. Consequently, the nozzle produces not only a reactionary force inherent to its design but also a secondary aerodynamic force, acting in the same direction and, though smaller, significantly impacting the overall behaviour.

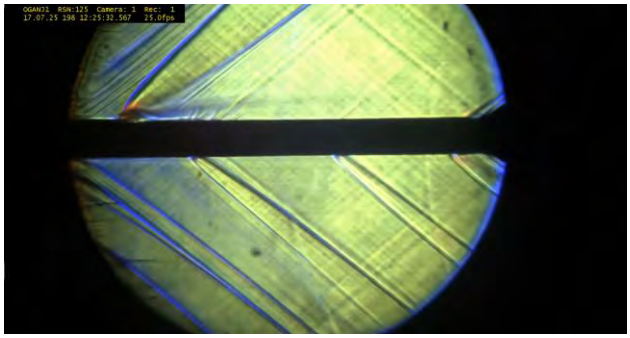


Figure 5. Schlieren system visualisation of shock waves generated by free stream and side jet interaction.

Schlieren system visualization provides a detailed view of the array of aerodynamic effects generated by the interaction. Visualization based on density differences provides an implicit view of changes in fluid properties – specifically, the shear layer, Figure 5. Rapid expansion of gas into the free stream field acts as a blunt obstacle against the free stream, resulting in the formation of a bow shock in front of the visible barrel shock. A small separation region is observed near the bow shock zone.

Shear layer vortices and more turbulent flow develop downstream and gradually recirculate into the main flow. Expansion of the side jet generated a Mach disc, along side the expected bowshock.

A heatmap overlay further revealed asymmetry in the pressure distribution, Figure 6. Subsequent analysis indicated that this asymmetry was a consequence of material discrepancies, which subsequently disrupted the already turbulent flow near the nozzle. In other words, the observed pressure asymmetry was primarily a result of the model's physical construction and had no direct relationship to the interaction of the side jet.

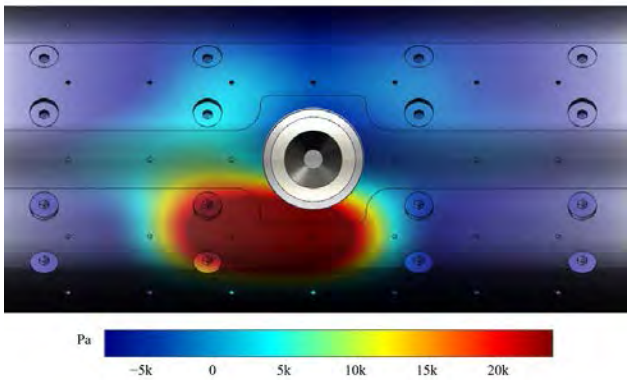


Figure 6. Pressure distribution over area near nozzle for supersonic crossflow (external flow *M*=1.6, side jet flow M=1)

With this data, it was possible to confidently deduce that the overall pressure distribution, under similar conditions, remained within a range of no more than 10 kPa.

Furthermore, Figure 6. provides in quantity and quality more broad view of general interaction between free stream and side jet. The stagnation zone can clearly be observed in front of the nozzle, along with zone of recirculation further downstream.

4. CONCLUSION

Effective control of aerodynamic bodies utilizing side propulsion systems depends not solely on the impulse generated by the reactive force, but also on a comprehensive understanding of the aerodynamic effects produced under these conditions.

The resulting bowshock, Mach disk, intensified turbulence, and flow separation further complicate the necessary control strategies.

Tests conducted under transonic conditions demonstrated that transonic flight generates the greatest aerodynamic contribution to the overall force, along with significant flow separation.

During these tests, it was clearly demonstrated that pressure distribution and flow separation, when considering the interaction between the freestream and the side jet, should not be dismissed. Indeed, careful consideration of flow separation is crucial in real flight conditions.

Acknowledgements

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-34/2026-03/200325).

References

[1] STAHL,B., ESCH,H., GULHAN, A.: *Experimental investigation of side jet interaction with supersonic cross flow,* Aerospace Science and Technology, 12 (2008) 269-275

[2] HUH,J.,LEE,S.: *Numerical study on lateral jet interaction in supersonic crossflow,* Aerospace Science and Technology, 80 (2018) 315-328

[3] STAHL,B., EMUNDS,H., GULHAN, A.: *Experimental investigation of hot and cold side jet interaction with a supersonic cross-flow,* Aerospace Science and Technology, 13 (2009) 488-496

[4] MEDVED, B., ELFSTROM, G.: *The Yugoslav 1.5m trisonic blowdown wind tunnel,* AIAA 86-0746, in: 14th Aerodynamic Testing Conference, American Institute of Aeronautics and Astronautics, 1986

[5] ILIĆ, B., OCOKOLJIĆ, G., SAMARDŽIĆ, M., ISKAOVIĆ, J.: *Upgrade of the VTI T-38 Blowdown Trisonic Tests Facility: Wind Tunnel Control System,* in: ICAS Proceedings -34th Congress of the International Council of the Aeronautical Sciences, International Council of the Aeronautical Sciences, Florence, 2024:p. ICAS2024_0456 (2024)

[6] SAMARDŽIĆ, M., VUJIĆ, N., ILIĆ, B., OCOKOLJIĆ, G.: *Experimental results from the missile model test campaign in the VTI T-38 wind tunnel*, 11thInternational Conference on Defensive Technologies, OTEH 2024 Proceedings, Tara, Serbia, 2024, pp.43-49.

[7] ILIĆ, B., SAMARDŽIĆ, M., OCOKOLJIĆ, G: *Integrated multi-technique experimental setup for wind tunnel investigation of side jet interaction with cross-flow*, Aerospace Science and Technology, 177 (2026) 112705.

Nomenclature

C_p	– Pressure coefficient
D	– Model diameter
d	– Throat diameter
L	– Model length
M	– Mach number
p	– Local surface pressure
p_∞	– Static pressure of the external flow
p_{ojet}	– Jet stagnation pressure
R_e	– Reynolds number
k	– Specific heat ratio for air
α	– Angle of attack

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION II

Aircraft - A

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SM2-M UAV AND Q GROUND CONTROL: SYNERGY OF AUTONOMOUS GUIDANCE AND REAL-TIME MISSION MONITORING IN MODERN MILITARY OPERATIONS

MILOŠ UGRENOVIĆ
Mixed Artillery Brigade, Niš, ugren.srbin@gmail.com

Abstract: *This paper analyzes the synergy between autonomous guidance and real-time mission tracking of the SM2-M unmanned aerial vehicle (UAV), with the aim of improving the reliability, accuracy and operational efficiency of the mission. The proposed system architecture integrates autonomous navigation, flight control and mission tracking algorithms with embedded sensor and communication subsystems, enabling continuous data exchange and real-time situational awareness. Special emphasis is placed on adaptive flight path management and decision-making in dynamic operational conditions. The analysis shows that an integrated approach to autonomous guidance and mission tracking improves situational awareness, reduces operator workload and increases mission success. The analysis was performed during flights with the SM2-M unmanned aerial vehicles that were implemented at the Pasuljanske livade test site. The results obtained indicate the potential for further development and application of advanced autonomous capabilities in the SM2-M unmanned aerial vehicle, contributing to the efficiency of next-generation military unmanned aerial vehicles.*

Keywords: *SM2-M UAV, autonomous guidance, mission monitoring, real-time systems, autonomous systems.*

1. INTRODUCTION

Unmanned aerial vehicles (UAVs), commonly known as drones, have become one of the most important tools in modern warfare. Thanks to advances in technology, they are used for reconnaissance, surveillance, intelligence gathering, artillery fire adjustment, electronic warfare, and carrying out precision strikes against military targets.

One of the main advantages of unmanned aerial vehicles is their ability to perform dangerous missions without putting pilots' lives at risk. Drones can remain airborne for extended periods, transmit real-time video and data, and operate in areas that are difficult to access or too dangerous for manned aircraft.

The ground Control Station (GCS) is one of the most important components of an unmanned aerial vehicle (UAV) system. It enables the operator to plan missions, control the aircraft, monitor its flight in real time, and supervise the operation of all critical onboard systems. Without a reliable ground control station, the safe and efficient operation of a UAV would be significantly more difficult. The function of the ground control station is to control the UAV and present all flight data to the operator. This communication allows the exchange of data such as the aircraft's position, speed, altitude, battery or fuel status, as well as information collected by onboard sensors and cameras. At the same time, the operator can transmit commands to modify the flight route, altitude.

In addition to controlling the aircraft, the ground control station plays a vital role in ensuring flight safety. The operator can promptly detect technical malfunctions, communication failures, or adverse weather conditions and take appropriate actions, such as modifying the flight plan or activating the automatic return-to-home function.

The importance of the ground control station becomes even more evident during complex operations, where multiple parameters must be monitored simultaneously and rapid decisions are required. For this reason, the development of modern ground control stations focuses on improving communication reliability, intuitive user interfaces, process automation, and protection against cybersecurity threats.

1. SM2-M VEHICLE MODEL

The SM2-M is a modern unmanned system that belongs to the category of loitering munitions. Unlike conventional reconnaissance unmanned aerial vehicles, whose primary purpose is to gather intelligence and return after completing their mission, the SM2-M (show in Fig. 1) is designed to locate a target, carry out an attack, and be expended during that single mission. This concept combines reconnaissance capabilities with precision strike functions in one integrated system. The system is designed as a lightweight modular platform, allowing for easier transportation and rapid deployment.

The aircraft is launched from the ground using a pneumatic launcher and is powered by a two-stroke internal combustion engine. It uses a Global Navigation Satellite System (GNSS) for navigation, enabling accurate guidance and the successful execution of assigned missions.

Corresponding author: M. Ugrenović;
E-mail: ugren.srbin@gmail.com
DOI: 10.5937/OTEH26008U

Figure 1. UAV on the launch pad

One of the most important features of the SM2-M is its ability to engage a variety of ground targets. The vehicle can be equipped with a warhead intended for attacking personnel and lightly armored vehicles, making it an effective means of conducting precision strikes. Before engaging its target, it can loiter over the operational area, observe the terrain includes a ground control station used for mission planning, flight control, communication with the vehicle, and pre-flight system checks, and identify the target, thereby increasing the likelihood of mission success.

The basic UAV physical characteristics are shown in Table 1 [1].

Table 1. UAV SM2-M physical characteristics

Parameters	Range
Drone length	2900 mm
Wingspan	4850 mm
Drone weight	75 kg
Flight autonomy	2 h
Communication range	23 km
Maximum speed	151 km/h
Engine displacement	178 cm³
Maximum tilt angle	40°

Component part of the system:

- drone CM2-M,
- remote controller,
- ground control station,
- transport vehicle [1].

2. GROUND CONTROL STATION

The ground control station (show in Fig. 2) is used to create a mission for the SM2-M unmanned aerial vehicle. Q Ground Control (QGC) is an open-source software application designed for mission planning, control, and monitoring of unmanned aerial vehicles (UAVs). It is developed as a universal ground control station (GCS) that supports a wide range of autopilot systems, with PX4 and ArduPilot being the most widely used. One of the primary functions of Q Ground Control is autonomous mission planning. Users can define flight paths by creating waypoints, specifying flight altitude, speed, mission actions, and other mission parameters. Once the flight plan is completed, it is uploaded to the UAV's autopilot, allowing the aircraft to execute the mission.

During flight operations, the application provides real-time telemetry monitoring. The operator can observe the UAV's position on a digital map, altitude, speed, heading, battery status, communication link quality, GPS signal strength, and many other parameters that are essential for safe mission execution. This real-time visualization enables rapid decision-making in unexpected situations. Autonomous guidance is one of the most important features of **Q Ground Control**, enabling an unmanned aerial vehicle (UAV) to execute a predefined mission without continuous operator control.

Figure 2: Ground Control Station

The basic Q Ground Control physical characteristics are shown in Table 2 [1].

Table 2. Q Ground Control physical characteristics

Characteristics	Description
Application type	Software for unmanned aerial vehicle control
License	Open-source software
Supported autopilots	PX 4, ArduPilot
Communication protocol	MAVLink
Flight monitoring	Real-time display of aircraft position, flight path, telemetry data
Navigation	Support for GPS/GNSS navigation
Mission planning	Creation of autonomous flight route using waypoint
Telemetry	Display of altitude, speed, heading, battery voltage, GPS status

The mission is planned (show in Fig. 3) within the application by creating a flight path consisting of a series of waypoints. For each waypoint, the user can define parameters such as flight altitude, speed, aircraft orientation, and specific actions, including taking photographs, hovering at a designated location, or returning to the launch point.

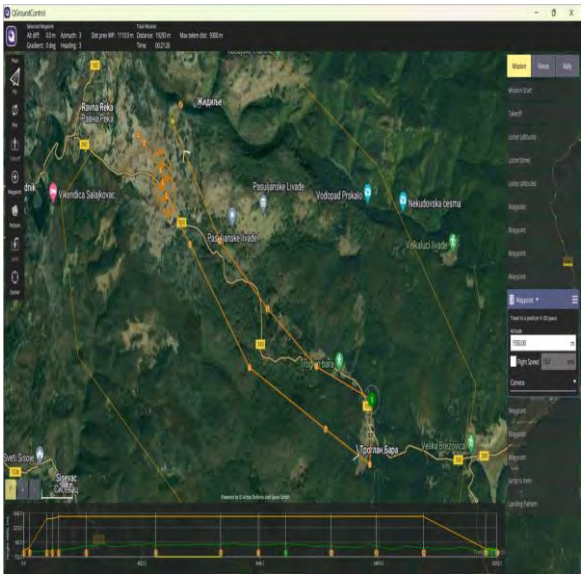


Figure 3. Mission planning

Once mission planning is completed, the flight plan is uploaded to the flight controller via the **MAVLink** communication protocol. The flight controller uses data from the Global Navigation Satellite System (GNSS) receiver, the Inertial Measurement Unit (IMU), magnetometer, barometer, and other onboard sensors to continuously determine the UAV's position and orientation. Based on this information, the autopilot continuously adjusts the motor outputs and control surfaces to ensure that the aircraft follows the planned flight path.

Q Ground Control also supports dynamic mission possible management during flight. The operator can add new waypoints, modify the existing route, adjust the flight altitude or speed, and switch between different autopilot flight modes without interrupting the mission. In the event of communication loss or other emergency situations, built-in safety features can be activated, such as **Return-to-Launch (RTL)**, automatic landing, or **Loiter** mode, depending on the system configuration.

During mission execution, Q Ground Control displays the position of the unmanned aerial vehicle on a digital map in real time. The operator can monitor the aircraft's current location, completed flight path, planned route, and the relationship between the actual and desired trajectory. This type of visualization allows deviations from the planned route to be detected quickly and enables timely responses if problems occur during the flight.

In addition to geographic position, the application displays a wide range of telemetry data continuously transmitted from the aircraft. The most important parameters include flight altitude, airspeed, heading, battery status, communication link quality, number of available GPS satellites, autopilot operating mode, and the status of individual aircraft systems. Based on this information, the operator can evaluate the current condition of the mission and make appropriate decisions.

One of the important capabilities of Q Ground Control is the display of data from sensors and cameras installed on the unmanned aerial vehicle. Depending on the system configuration, the operator can monitor live video transmission, analyze collected data, and associate it with the current position of the aircraft. This function is particularly important for tasks such as terrain surveillance, infrastructure inspection, mapping, and research missions.

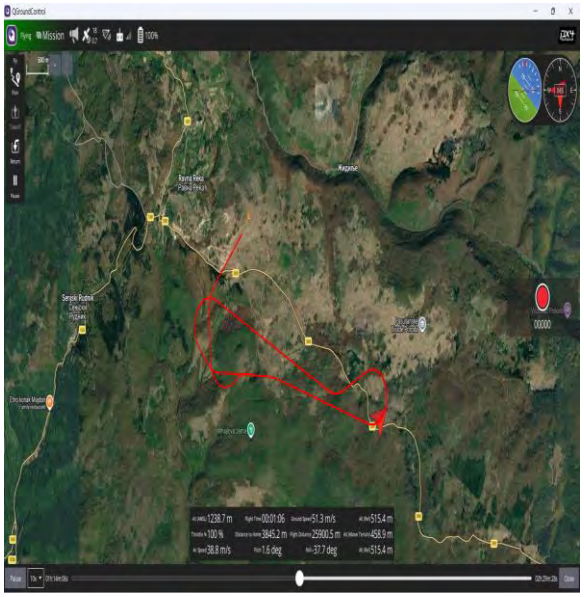


Figure 4. Real-time mission monitoring

Real-time mission monitoring (show in Fig. 4) also contributes significantly to flight safety. Q Ground Control provides warnings about potential issues such as low battery level, loss of GPS signal, communication failure, or exceeding predefined flight limits. In emergency situations, the operator can modify the mission plan, take manual control, or activate safety procedures such as an automatic return to the launch location.

During the execution of a mission that integrates autonomous drone guidance with a real-time mission tracking system, the role of the operator has significantly changed from that of traditional unmanned aerial vehicle (UAV) control. Instead of continuously piloting the drone, the operator has taken on the role of a mission supervisor, monitoring the operation of the autonomous system, analyzing the collected data, and making decisions whenever necessary. This approach allows for a more efficient distribution of tasks between the human operator and the autonomous system.

The operator at the ground control station received the following data:

- Flight mode: Mission
- Number of satellites: 18
- Battery level: 100%
- Flight altitude: 1238.7 m
- Throttle percentage: 100%
- Airspeed: 38.8 m/s
- Distance from launch point: 3845.2 m
- Pitch: 1,6 deg
- Flight distance: 25900,5 m
- Relative height: 515,4 m

The operator's primary responsibility was to continuously monitor telemetry data, including the drone's position, altitude, flight speed, battery status, communication link quality, and sensor status. At the same time, the operator viewed and other data collected during the mission, providing a comprehensive overview of the mission's progress and allowing for an assessment of whether the mission was being executed according to plan. An especially important aspect of the operator's work is the analysis of information generated in real time by the autonomous system and artificial intelligence algorithms.

Automatically detected objects, obstacles, or potential hazards are presented to the operator, who evaluates their relevance and, when necessary, decides whether to modify the flight plan, or perform a more detailed inspection of a specific area. In this way, human judgment complements the capabilities of the autonomous system.

In addition to monitoring, the operator retains the ability to intervene at any stage of the mission. In the event of changing weather conditions, loss of communication, reduced battery capacity, or the appearance of unexpected obstacles, the operator can modify the flight route, assign a new mission target, initiate the drone's return-to-home procedure, or take over manual control.

A ground station operator should have knowledge of satellite communications and radio systems, antenna systems and tracking equipment, TT&C & (Telemetry, Tracking and Command) operations, the fundamentals of flight dynamics, emergency procedures and operational risk management.

5.CONCLUSION

The synergy between autonomous onboard decision-making and ground-based supervision allows drones to adapt to dynamic environments while maintaining communication with operators and mission control systems. Autonomous algorithms can respond instantly to changing flight conditions, whereas real-time monitoring provides situational awareness, verifies mission execution, and enables timely human intervention whenever necessary. This balance between autonomy and human oversight significantly reduces operational risks, improves mission reliability, and increases the overall resilience of UAV systems.

As autonomous technologies continue to evolve, supported by advances in artificial intelligence, sensor fusion, communication networks, and edge computing, the interaction between onboard autonomy and real-time mission monitoring will become even more sophisticated. Future UAV systems are expected to operate with greater independence while remaining fully integrated into supervisory control architectures that ensure transparency, accountability, and operational safety.

References

[1] SM2-M.: *User Manual*, 19

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AIRCRAFT LIFE EXTENSION THROUGH INTEGRATED STRUCTURAL ASSESSMENT AND SYSTEM REVALIDATION

MARINA OSTOJIĆ
UTVA Aircraft Industry, Pančevo, majce74@yahoo.com

TAMARA NIKOLIĆ
UTVA Aircraft Industry, Pančevo, tn.aircraft@yahoo.com

MARKO PROŠIĆ
UTVA Aircraft Industry, Pančevo, proskezmn@gmail.com

Abstract: *Aircraft overhaul is a process of thorough restoration and upgrade, with a purpose of tackling exploitation and time-affected wear, eventually resulting in significant service life extension, update of the existing, and integration of additional, modern, features, while enhancing reliability and maintaining compliance with regulatory requirements. Timely overhaul is essential for ensuring safety, reliability and airworthiness of a long serving aircraft, and keeping it's systems up-to-date and compatible with new regulations. It begins with detailed inspection, accompanied by disassembly and refurbishment of major components wherever appears necessary. Inspection includes visual checks, non-destructive testing (NDT), and electronic diagnostic procedures to detect defects, wear or damage, particulary structural, mostly omitted in periodic checks. The components that undergo overhauls include engines, propellers, landing gear, avionics, built-in systems, control surfaces and aircraft structure. After the repair work is complete, overhauled aircraft undergoes rigorous testing and quality assurance checks to verify its performance, functionality, and airworthiness.*

In this work, a comprehensive overhaul process of Lasta aircraft will be shown as an example of such an undertaking.

Keywords: *Overhaul, inspection, NDT, defects, QA checks.*

1. INTRODUCTION

Serial Lasta 3 (military designation V-54) highly manoeuvrable light trainer aircraft were produced by Utva Aircraft Industry and delivered to Serbian airforce by 2015. Having flown extensively ever since, some are close to or have already reached its operational limit period and are in need of detailed examination and evaluation of current condition to continue safe operation.

One of these aircraft with S/N 54204 has recently undergone a thorough examination and repair work i.e. overhaul within Utva facilities.

This document will present contents and volume of work that overhaul requires, especially inspections with their findings and test results that had been needed to obtain in order to announce the airplane airworthy, shown on the sample with focus on NDT inspection, corrosion tackling and flight controls proper operation.

2. THE OVERHAUL PROCESS

The overhaul process is divided in stages, each with typical demands and scope of work. These stages and brief display of its contents are listed in Table 1.

Table 1. Overhaul stages and contents

1. Aircraft reception
Current state inspection; Antennae & sensors removal; Airplane transport to overhaul facility
2. Reception inspection, cleaning, scheduled maintenance
Disassembly, cleaning and inspection (DCI) of wings, stabilizers, controls & control surfaces, electrical wiring, fuel system, hydraulic system, pitot system, air conditioning system, engine, engine mount, avionics
3. Repair work, global & components
Wing, fuselage & empennage structure NDT and repair; Fuel tank integrity; Corrosion removal & protective coatings re-application; Windshield, windows; Interior equipment, furniture; Actuators; Instruments & electronics testing, calibration;
4. Repair work, outsourced
Engine; Propeller; Landing gear; Engine mount, wing & other main mounts radiography; Special protective coatings

Corresponding author: M. Ostojić;
E-mail: majce74@yahoo.com
DOI: 10.5937/OTEH26009P

5. Components assembly
Wing completion, fuel, hydraulic, electric and pitot systems assembly; Flight controls completion; Fuselage completion, fuel, hydraulic, electric and pitot systems assembly; Flight controls completion; Secondary controls installation; Control panel & avionics installation;
6. Airplane assembly
Wing installation; Main landing gear installation; Control surfaces installation; Engine mount installation; Nose landing gear installation; Fairings installation; Engine, propeller & exhaust installation; Interior equipment & canopy installation;
7. Completion and final tests
Painting and marking; Jacking and levelling; Geometry, weight measurement, centre of gravity check; Lubrication; Flight controls adjustment; Pitot system final test; Fuel system test & instruments calibration; Hydraulic system refill & Landing gear test; Final engine and engine control system tests; Air conditioning system tests; Antennae and sensors installation; Final electric & avionics tests; Engine de-conservation, run-up & propeller dynamic balancing; Engine controls adjustment; Detonation cord installation
8. Transportation and flight testing preparation
Compass calibration; Aircraft preparation, engine start-up; Engine parameter monitoring; Pre-flight check

While stages might overlap, depending on progress being made and pace at which separate jobs are being done, there are certain dependencies between many of them, preventing convenient deviations off the prescribed order of work most of the times.

Chosen part of overhaul process to be presented in details further in this work includes NDT inspection, corrosion reparation and flight controls test.

2.1. Non-destructive testing - NDT

Non-destructive testing methods were performed on various aircraft structural components, including fittings of the horizontal stabilizer, vertical stabilizer, fuselage, left and right wings, also primary flight control components and so on, in order to assess the condition of these vital parts.

The following NDT methods were applied:

- Visual Inspection (VI)
- Penetrant Inspection (PI)
- Eddy Current Testing (ECT)
- Ultrasonic Testing (UI)

Visual inspection

Inspections by means of close and detailed observation, aided or not with convenient tools, optical devices, measuring equipment including go/no go and angular fixing tools etc. make for VI methods. These methods are the most common and first to be used to notice obvious and measurable defects, being convenient to use, economic and always on-hand. Its main limitation is low sensitivity and inability to detect sub-surface and micro level damage.

Liquid Penetrant Inspection

Liquid penetrant inspection is a non-destructive method used to detect surface-breaking discontinuities in non-porous materials. It is suitable for detecting cracks, laps, porosity, shrinkage cavities, laminations, and similar discontinuities. It is simple, cost-effective, and capable of detecting very small surface cracks; however, it is limited to surface-breaking defects only.

Ultrasonic Inspection

Ultrasonic testing uses high-frequency sound waves to detect both surface and subsurface discontinuities. It is widely used for detecting cracks, laminations, voids, bonding defects, and corrosion. It provides high accuracy in determining defect location, size, and orientation.

Eddy Current Inspection

Eddy current testing is based on electromagnetic induction and is used for detecting surface and near-surface defects in electrically conductive materials.

It is commonly used for detection of fatigue cracks and stress corrosion cracking around fastener holes. The method does not require direct contact and provides fast inspection capability, but penetration depth is limited.

2.2. NDT methods used

• VI was performed on all fittings using 5× magnification lenses and a borescope. A borescope inspection was carried out in all inaccessible areas to detect hidden defects and corrosion.

• PI was performed on primary control fittings, bushings, wing front and rear attachment fittings, and spars.

• ECI was performed on all metallic and non-metallic conductive fittings to detect cracks and surface defects.

• UI was performed on all structural joints and fittings.

Aircraft inspection and repair

Proper identification of aircraft structural materials is essential to ensure continued airworthiness and to prevent improper repairs.

Effective structural repair design requires understanding of applied stresses, including tension, compression, bending, torsion, shear, and bearing.

Aircraft repairs must balance structural strength and weight efficiency. Repairs must provide adequate load-carrying capability with the required safety factor, while avoiding excessive stiffness that may introduce stress concentrations and secondary cracking.

2.3. NDT detected defects

A couple of examples of inspection findings are listed in the Table 2. Based on the results obtained by the applied inspection methods, it can be concluded that no unacceptable defects were found on fuselage fittings, left and right wing fittings, vertical and horizontal stabilizer fittings, and aircraft flight control components.

Table 2. Defect examples

1. Horizontal stabilizer to elevator joint hinge fitting	
Methods used: VI, ECI, UI	Findings: Mechanical damage (Ø4 mm, Ø5 mm, Ø7 mm) – **Figure 1**
2. Bellcrank, elevator control	
Methods used: VI, PI, ECI, UI	Findings: Mechanical damage, L = 10 mm, D = 0.3 mm
3. Bellcrank, rudder control	
Methods used: VI, PI, ECI, UI	Finding: Mechanical damage, L = 15 mm, D = 0.5 mm
4. Cylinder fitting, right wing	
Methods used: VI, ECI, UI	Finding: Damage of anodized coating (B+S7)

Figure 1. Example of VI noticeable mechanical damage

Some of typical defect mostly sought after at overhaul are signs of fatigue, corrosion and cracks. Here they are going to be explained more closely, particularly referring to EASA regulations.

2.4 Fatigue

The objective of fatigue related requirements is to ensure the retention of both limit and ultimate load capability throughout the service life of the aircraft.

For wings, empennage, and associated structures, compliance may be demonstrated by comparison with a similar, service-proven design with equivalent stress levels, materials, stress concentrations, and operational usage.

For small aircraft certification categories, these requirements apply to the pressurized cabin, wings, empennage, and associated structures. Landing gear and engine mounts are excluded.

2.5 Corrosion

Metals corrode due to chemical or electrochemical interaction with the environment. Corrosion worsens significantly by salt contamination, moisture presence, dirt deposits and so on. Alclad alloy 2024 used for aircraft skins is protected by a pure aluminium layer, surface protection systems, and paint coatings. At dissimilar metal interfaces (e.g., steel-to-aluminium), contact surfaces are protected with PS 892 C12 sealant.

Three criteria are used to determine whether a corroded component may be repaired:

- Component classification.
- Repair category.
- Corrosion severity.

The most common types of corrosion observed during aircraft service life are:

- General corrosion - most frequently found on steel and ferrous components, it appears in the form of reddish rust. Since it is easily detected by VI, it is generally not considered a hazardous form of corrosion. It usually develops where protective coatings have been damaged, particularly on cadmium-plated steel components.
- Pitting corrosion - the most common form of corrosion on polished aluminium surfaces. It appears as white or grey powdery deposits. After removal of the deposits, small pits become visible on the surface.

Pitting corrosion develops in three stages: pit initiation, early pit growth and stable pit growth. It is normally detected during the second stage. The coalescence of several pits may result in crater formation.

- Galvanic corrosion – it occurs when two dissimilar metals are in electrical contact in the presence of an electrolyte.

A typical example is the interface between an aluminium alloy component and a steel fastener. The greater the difference in electrochemical potential between the materials, the faster the corrosion process. For example, aluminium in contact with Monel may experience rapid and severe corrosion.

- Intergranular corrosion – the type that develops within the material structure and attacks the internal microstructure. It is often not detected until severe deterioration has already occurred.

Rapid intergranular may occur in 2024 and 7075 alloys if improper heat treatment has been applied.

Although material loss may be minimal, mechanical properties are severely degraded. The material becomes brittle and may fail under relatively low static or dynamic loads.

Exfoliation corrosion is an advanced form of intergranular corrosion characterized by separation of material layers.

- Crevice corrosion occurs at metal-to-metal interfaces, particularly at the edges of joints or between closely fitted surfaces.
- Filiform corrosion is a special form of crevice corrosion occurring beneath organic protective coatings.

It is recognized by thread-like patterns beneath the coating surface. This type of corrosion develops at relative humidity levels between 78% and 90% and temperatures above 30°C.

Repair categories

1st category

1.1 Components identified during corrosion inspections for which material removal during repair will not reduce the thickness below the minimum allowable thickness.

1.2 Components whose damage exceeds the minimum allowable thickness and whose repair is non-standard and not commonly encountered on other aircraft of the same series. Such repairs shall be specifically defined and approved by the manufacturer.

1.3 Components that repeatedly exhibit minor corrosion during service and are identified during every corrosion inspection. If it is determined during the latest inspection that corrosion removal would reduce the remaining thickness below the minimum allowable value, repair may still be permitted provided that the repair is specifically defined and approved by the manufacturer.

2nd category

Includes components corresponding to Class 1 and Group 1.3; however, these components shall be reinforced during repair.

3rd category

Includes components for which initial or subsequent corrosion inspections reveal severe corrosion levels that may present a significant hazard to aircraft operation.

Corrosion Severity

Three levels of corrosion severity are defined:

Level 1 – Surface Corrosion

Surface corrosion affecting material to a depth of up to approximately 0.0254 mm.

Level 2 – Moderate Corrosion

Corrosion affecting between approximately 0.0254 mm and 10% of the material thickness.

Level 3 – Severe Corrosion

Corrosion affecting more than 10% of the component thickness and potentially reaching up to 15% of the material thickness.

Repair of corroded components

Several standard methods are used for corrosion removal, including chemical treatment, manual sanding using aluminium oxide abrasive paper or metallic wool made of material compatible with the surface being treated, and mechanical sanding or polishing using abrasive wheels or pads.

During the inspection of Lasta S/N 54204, the airplane was found to be remarkably clear of corrosion, requiring only minimum intervention and cleaning actions. This is a testimony of surface protection well done during manufacture and assembly, combined with proper greasing and lubrication during service and timely periodic maintenance.

3. FLIGHT CONTROLS INSPECTION

After the airplane assembly stage is complete, it is time for completion and final tests stage. One of the most significant tasks in this stage is flight controls functional test and adjustment. There are multiple requirements that controls and control surfaces have to meet to acquire final approval, most important of which being proper deflections, proper input to output force ratio, acceptable free play and low friction. In this chapter, results of some of these tests are displayed that quantify proper state of controls after the assembly.

Results of measurements of flight control system free play, static friction and control forces on aircraft v54204

Control stick - elevator free play: 1,2 mm (maximum permissible value 2 mm) (**Figure 2**)

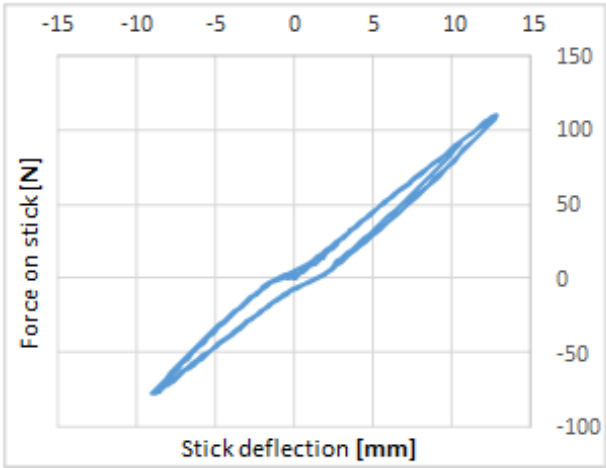


Figure 2. Control stick – elevator, free play

Elevator control system static friction force: 0,66 daN (maximum pesrmissible value 1 daN) (**Figure 3**)

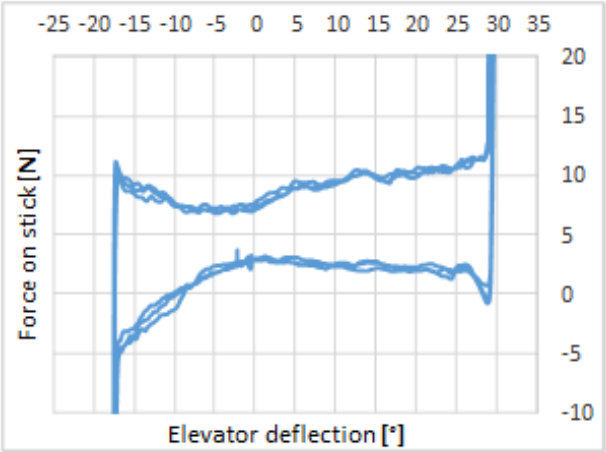


Figure 3. Elevator control static friction

Elevator control force with the elevator trim in the neutral position:
- Forward stick: 58 N
(reference value 62±5 N)
- Stick in the neutral position: +8 N / -8 N
(reference value 8±2 N)
- Aft stick: 96 N
(reference value 95±5 N) (**Figure 4**)

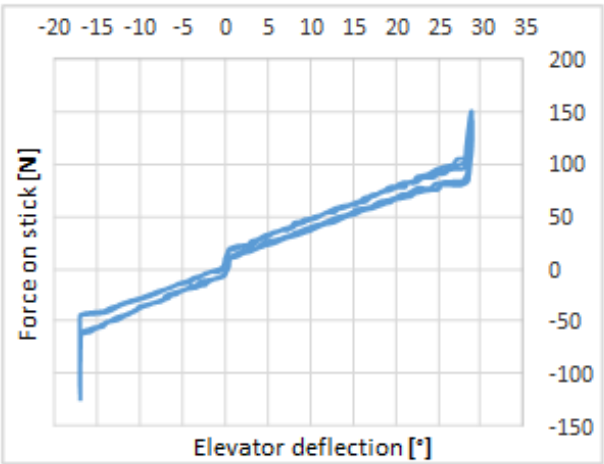


Figure 4. Elevator control force, no trim

Elevator control force with the elevator trim in the neutral position:
- Forward stick: 58 N
(reference value 62±5 N)
- Stick in the neutral position: +8 N / -8 N
(reference value 8±2 N)
- Aft stick: 96 N
(reference value 95±5 N) (**Figure 5**)

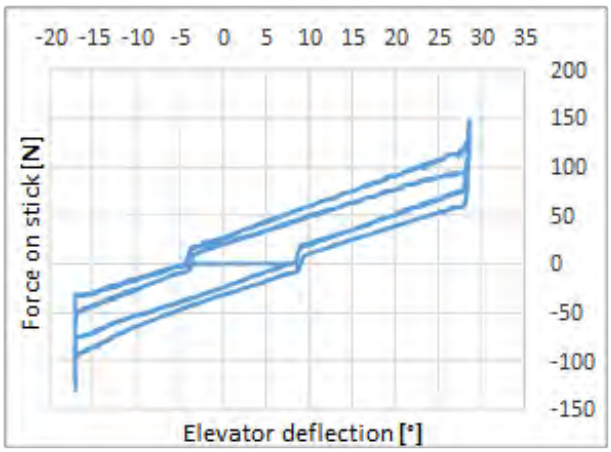


Figure 5. Elevator control stick force, maximum trim

Left aileron control stick free play: 1,5 mm (maximum permissible value 2 mm) (**Figure 6**)

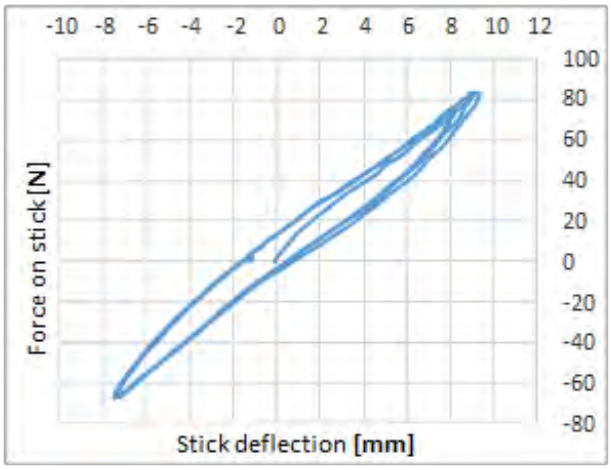


Figure 6. Control stick – left aileron, free play

Right aileron control stick free play: 0,8 mm (maximum permissible value 2 mm) (**Figure 7**)

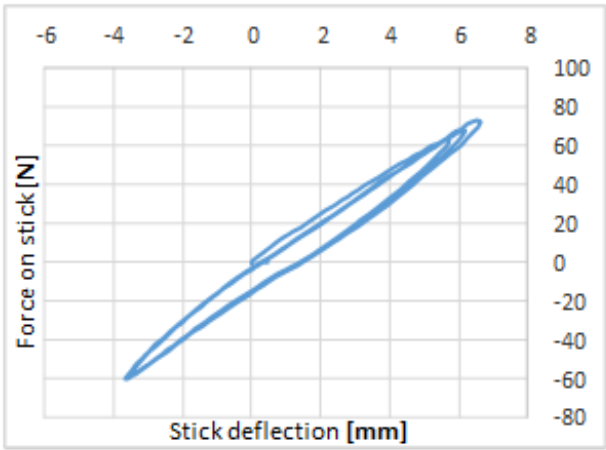


Figure 7. Control stick – right aileron, free play

Aileron control system static friction force: 0,48 daN (maximum permissible value 1 daN) **(Figure 8)**

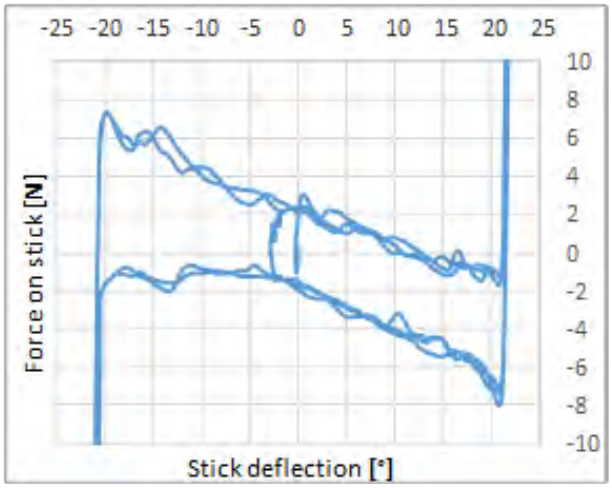


Figure 8. Aileron control static friction

Rudder control system static friction force: 1,8 daN (maximum permissible value 2,5 daN) (**Figure 9**)

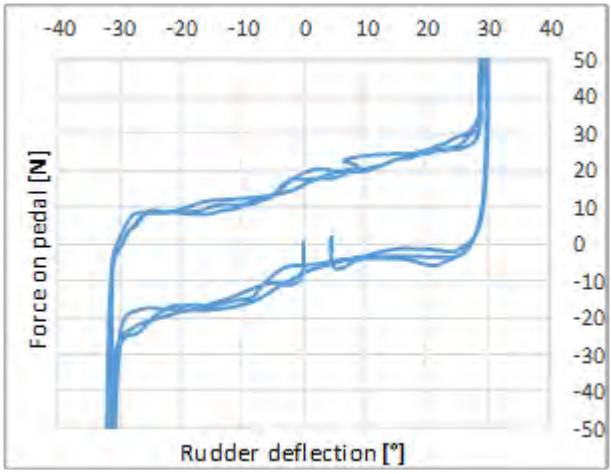


Figure 9. Rudder control static friction

Rudder pedal free play: 1,9 mm (maximum permissible value 2 mm) (**Figure 10**)

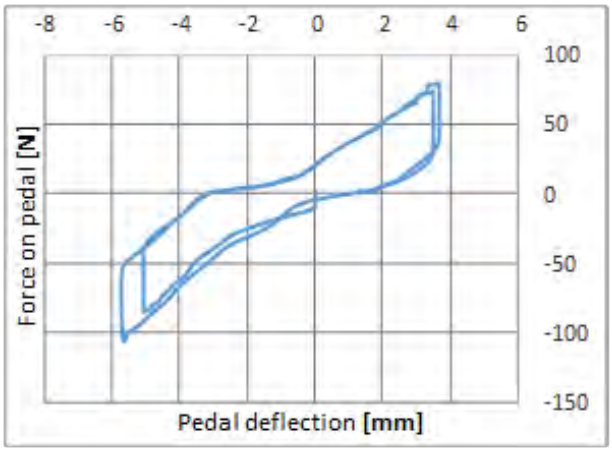


Figure 10. Rudder control, free play

4. CONCLUSION

During the overhaul process, Lasta 3 (V-54) airplane with S/N 54204 has been inspected, refurbished and tested. After successful completion of factory and flight tests, its airframe has been announced airworthy for a period of at least another 6 years or 1000 flight hours, whichever expires first, with periodic inspections and maintenance carried out as prescribed in AMM.

While overhaul goal is 10 years airframe service life extension, it is constrained to 6 years with an expiry date set by design at the time of production of the airplane S/N 54204 to be the year 2032.

Individual components' service life is subject to periodic maintenance, and can be extended depending on condition far beyond airframe service life expiry.

Conducting airplane overhaul and repair work on demand positions Utva Aviation Industry as a valuable MoD asset that continues successful work in maintaining Serbian air force's fleet of light trainer aircraft up and running for new generations of cadet pilots to improve and master their flying skills before advancing to jet fighters.

References

[1] AC 43-4A/4B - Corrosion Control for Aircraft

[2] AC 23-13A - Fatigue, Fail-Safe, and Damage Tolerance Evaluation of Metallic Structure for Normal, Utility, Acrobatic, and Commuter Category Airplanes

[3] AC 43.13-1B Change 1 - Acceptable Methods, Techniques, and Practices-Aircraft Inspection and Repair

[4] Certification Specifications for Normal, Utility, Aerobatic and Commuter Category Aeroplanes CS 23, Amendment 3 Book 1

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DRONE MILITARY TESTING

STEVAN JOVIČIĆ
Technical Test Center, Belgrade, msmjovicic@gmail.com

DANILO VUKOSAVLJEVIĆ
Technical Test Center, Belgrade, rmk997@live.com

NIKOLA ZDRAVKOVIĆ
Technical Test Center, Belgrade, nzdravkovicn@gmail.com

JELENA ŠOBOT
Technical Test Center, Belgrade, jsobot@outlook.com

Abstract: *Drone military testing include a wide range of experiences, from technical challenges to practical problems in the field. Here is what it usually looks like in practice: Technical tests checks how the drone functions externally: Field Experiences Real-world and real-time conditions are often the greatest challenge: Wind can significantly affect small drones; Temperature, cold quickly drains the battery; Signal interference, power lines, and carbon-metallic constructions; Pilot visibility loss of visual contact increases risk. In practice, testers often say that an „ideal laboratory result is rarely achievable outdoors". Our, Serbian army way of testing is standard military drone certifications. A critical standard testing resistance to extreme temperatures, mechanical shock, vibration. EMI/EMC Compliance: Ensures the drone's electronics do not interfere with other friendly military communication or radar systems. What is the quality of drones made in the Republic of Serbia from domestic development, state projects and private manufacturers is presented in this paper.*

Keywords: *drone, examination, rule, swarm drone*

1. INTRODUCTION

Drones have become the hit of limited-intensity conflict in the early 21st century. War has become a game again, and drones are the enforcers of punitive measures. However, the main drawback of drones as a perfect weapon for now is its limitations in terms of the payload it can carry and the possibility of electronic jamming. When these issues are resolved, we will be able to talk about weapons that can bring victory in the war and create the conditions for the penetration of larger army groups on the front line. For now, we can say that victory in the war cannot be achieved only through use of drones.

However, even this previous claim became questionable during the conflict in Ukraine. The latest generation of drones has enabled a nearly permanent blockade of logistic supply routes used to resupply units, not only those on the front line but also throughout its strategic depth.

In this way, drones have become weapons that determine the outcome of armed conflicts, and as such, their quality must be verified before they are introduced into operational use.

Our, Serbian army way of testing is standard military drone certifications. A critical standard testing resistance to extreme temperatures, mechanical shock, vibration influence testing and EMI/EMC compliance.

2. DRONE TESTING RULES

According to Article 85 of the Rulebook on equipping the armed forces with weapons and military equipment, the procurement of weapons and military equipment as a finished product from the domestic or foreign market is carried out when a market procurement model is envisioned by a previous analysis, an Urgent equipping elaborate, or an Implementation program.

In the process of procuring weapons and military equipment from the market, the contract with the supplier must provide for the testing of the equipment, which is to be conducted during the acceptance of the equipment by a commission formed by the decision of the tactical holder.

The results of the verification testing of samples, or the testing results obtained during factory acceptance at the manufacturer's facility or final acceptance of the equipment, form the basis for the adoption of the equipment into the inventory of the Serbian Armed Forces.

The testing of military drones includes a series of rigorous technical, tactical and combat tests to determine their operational readiness, resistance to electronic interference and precision in destroying targets. In the modern context, the focus of the testing has shifted to flight autonomy driven by artificial intelligence, integration of drone swarm and action in conditions of active anti-aircraft defense [1].

Corresponding author: S. Jovičić
E-mail: msmjovicic@gmail.com

DOI: 10.5937/OTEH26010J

The examination covers several areas such as:

- Software and laboratory simulations: Verification of flight algorithms, resistance examination to electronic interference and operation of navigation systems.
- Climate-mechanical tests in the form of checking resistance to environmental influences.
- Operational service trials: Flights on military training grounds to test payload, durability, thermal imaging camera characteristics and operational range.
- Tactical evaluation: Examination of accuracy in hitting targets.

3. DRONE TESSTING EXPIRIENCE

Upon initial inspection of the assets received for evaluation, it is immediately apparent that drone manufacturing relies predominantly on components sourced from the Far East, specifically Asia, China, South Korea, and Taiwan.

Not only are the components sourced from the Far East, but every single software package shares that same origin. After an unusually prolonged development period, the domestic „Pegaz" project finally had to go and „"borrow brains" from Chinese partners. Operational service trials of military drones represent a key phase in which unmanned aerial vehicles are tested in realistic atmospheric, geographical and electronic conditions at dedicated military training ground.

These tests aim to verify the tactical-technical characteristics of the drone before it enters serial production or operational use.

- Flight performance check: Testing of maximum speed, flight ceiling, operating autonomy and behavior of the aircraft in extreme wind gusts, figure 1.

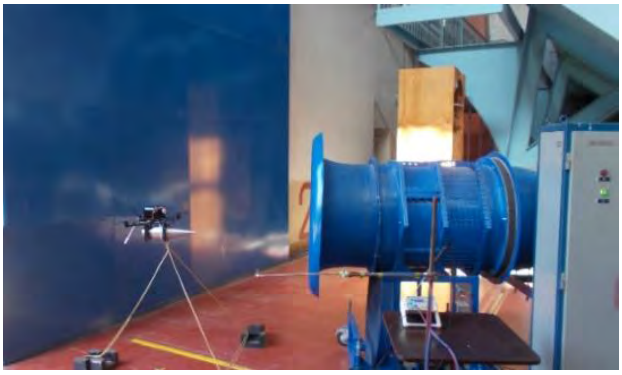


Figure 1. Testing of wind influence on drone flight characteristics

- Payload and stability testing: Monitoring the stability of a drone with a maximum payload, including electro-optical reconnaissance systems or ordnance.
- Immunity to electronic jamming: Simulate enemy counter-electronic (EW) activities to verify the stability of satellite navigation and radio communications.
- Optoelectronic equipment testing: Checking the range and accuracy of thermal imaging cameras, laser rangefinders and automatic target tracking systems.
- Maneuvering under load: Performing sharp maneuvers and simulating anti-aircraft defense (ADF) evasion.
- Combat shooting: Testing the accuracy of launching rockets, guided bombs or activating a kamikaze drone against static and moving targets.
- Take-off and Landing Systems: Checking the reliability of catapult, vertical take-off (VTOL) or parachute systems for forced descent.

Current domestic project

The „Komarac" drone

FPV kamikaze drone and interceptor family of drones represents the most massive domestic project in the segment of stray ammunition with following technical characteristics: Quadcopter with an operational range of 10 to 15 kilometers and a payload capacity of up to 5 kilograms.

Figure 2. „Komarac 1" drone, general appearance

Armament: The „Komarac 1" version carries a 64 mm cumulative warhead from the M80 „Zolja" hand-held rocket launcher, capable of penetrating up to 300 mm of homogeneous steel armor.

The latest development (Komarac 2V): During the last military demonstrations, Komarac 2V, a specialized disposable interceptor drone intended for the kinetic destruction of other enemy drones and missiles in the air, was presented for the first time.

„Pegasus" -reconnaissance-combat aircraft

„Pegasus" is the largest and most complex domestic unmanned aerial vehicle project, developed in cooperation with partners from China to integrate advanced data transmission and fire control systems.

Figure 3. „Pegasus" drone during take-off

Role in the field: It is intended for deep reconnaissance, observation, marking targets with a laser beam and carrying out attacks with guided air bombs and missiles.

Tests: The aircraft went through a long cycle of polygonal checks of the stability of the connection in the presence of electronic jamming and the precision of laser guidance on polygonal targets, figure 3.

„Vrabac"-tactical scout

The „Vrabac" is a smaller, easily portable aircraft that has been in operational use by the Serbian Army for years and serves as a basic tool for day and night reconnaissance at close distances, figure 4.

Figure 4. „Vrabac" drone in flight

Operational application: It is launched by hand or by means of an elastic rope (catapult), and it lands with a parachute.

Improvements: Its modified versions with more advanced electro-optical systems (thermal imaging cameras) and software for automatic recognition of targets on the ground are regularly tested on training grounds.

In addition to the mentioned, other domestic systems are increasingly appearing in military tests:

IKA (Bomber) drone

Domestically produced multicopters that are being tested for precise dropping of mortar mines on enemy positions. A product of the private company PRDC that has passed testing at TOC. It showed its quality at the recent demonstration at the „Pasuljanske livade" training ground, figure 5.

Figure 5. IKA bomber drone in flight

VTOL „Shadow":

Shadow is the VTI's latest project with vertical take-off and landing, which combines the speed of a classic airplane and the flexibility of a helicopter, figure 6.

Figure 6. „Shadow" drone, general appearance

Various producers offer various types of Tactical Drones and Military Drones for emergency management and mission-critical operations. These rugged military drones are designed to meet the demanding needs of Border Security and Armed Forces and are optimal for a range of applications such as high-altitude logistics delivery, ISR, mine detection and neutralization, combat, and target decoys, among others. On the market can be found customized tactical drones based on user-specific requirements, to meet the varying demands of weight, flight altitude, payload capability, endurance, armament and ammunition integration, electro-optical features and ruggedness.

The advanced AI technologies and visual algorithms in tactical drones and tactical military drones aid intelligent mapping of defined geographical areas; detect and classify objects of interest at low light and harsh weather conditions. These cutting-edge features make tactical military drones ideal for cross-border surveillance and to gather crucial situational awareness during a conflict and gather intelligence on border activities.

The rugged tactical drones are designed to transport in small SUVs and trucks and can be deployed in minutes from any terrain. These tactical military drones have a varying operational range of 3-10 km and can fly at an altitude of 1.500 m to 6.800 m, based on the payload, electro-optical sensor specifications.

Tactical military drones are equipped with advanced technology that enables them to perform a variety of

functions, including surveillance, reconnaissance, target acquisition, and even weapon delivery.

The tactical drones usually come with three flight modes, Manual, fully Autonomous and pre-mapped Autonomous. The user can define the mode of flight based on operational needs in these rugged Military Drones. For instance, for periodic cross-border surveillance at a defined area, a fully autonomous mode would be ideal. However, during a conflict wherein every second is crucial, Manual or semi-autonomous operation would provide quick insights to the operator. In case of link loss (loss of connectivity with GCS) due to signal jamming or extreme weather conditions, the Auto Return Home feature aids the drones to return to the launch safely. Producers also offers Tactical Military drones for distant surveillance with a range of up to 45 km, especially with stealth capabilities.

Rugged tactical drones come with ergonomically designed, user-friendly Ground Control Stations (GCS). The multi-layered safety protocols ensure secure communication between GCS and the drones. Based on the application and use case, the GCS can be designed to have single or dual displays to provide maximum usability and operational convenience to the operator on the field.

5. NEW THREAT FROM DRONES

China has for the first time publicly demonstrated the complete operation process of the Atlas system for operations with swarm drone units, thereby further emphasizing the accelerated development of autonomous military technologies. The system was presented through testing in real-world conditions, where drones carried out coordinated reconnaissance and precise attack on the target.

Figure 7. „Atlas" drone swarm launcher

The Atlas system consists of several components, including the Swarm-2 combat vehicle, command vehicle and logistical support. One Swarm-2 vehicle can launch up to 48 drones, while the command vehicle simultaneously manages up to 96 units in the row. During the demonstration, the system independently performed target recognition, drone launch and precise attack, without direct human control over each individual device. Drones in flight coordinated movement, avoided collisions and adjusted the formation in real time. The system's special strength lies in so-called "royal intelligence", where each drone functions as part of a wider network.

Thanks to advanced algorithms and AI processing, units exchange data, adjust positions and optimize the tension and output tasks [3].

The conflicts of 2025 proved that the electromagnetic spectrum is no longer a secondary domain. It is a primary arena of competition. Jamming, spoofing, cyber intrusions and attacks on satellite infrastructure changed the tactical landscape more than artillery or armor in multiple theaters.

This means one thing. Systems deployed in 2026 must be capable of operating through degradation. They must navigate without GPS, communicate across multiple resilient channels, and continue missions even under heavy interference. This reshapes how platforms are designed, procured and evaluated [4].

Producers offer Tactical military drones in a range of sizes and capabilities, from small quadcopters to large Octocopters or UAV plane shape drone.

The tactical drones usually come with three flight modes, Manual, fully Autonomous and pre-mapped Autonomous. The user can define the mode of flight based on operational needs.

6. FUTURE TESSTING

Coordination of drone swarm: Testing the ability of mutual communication and synchronized attack of multiple aircraft simultaneously without human assistance.

The past decade has seen an enormous increase in the use and importance of the "electronic battlefield." As a result of the continuing growth in this area, this volume has been prepared to serve as an introductory reference to the process of testing and evaluating Electronic Warfare Systems. Electronic Warfare (EW) is the mission area responsible for establishing and maintaining a favorable position in the electromagnetic domain.

Test and evaluation (T&E) of those devices used on modern military aircraft to prosecute this critical mission area requires the use of a wide range of test techniques and analytical methods to assure users of the readiness of EW systems to meet the challenges of the combat environment. Actual in-flight testing comprises a relatively small portion of the EW T&E process [5].

7. CONCLUSION

Drones have become the hit of limited-intensity conflict in the early 21st century. War has become a game again, and drones are the enforcers of punitive measures. However, the main drawback of drones as a perfect weapon for now is its limitations in terms of the payload it can carry and the possibility of electronic jamming.

When these issues are resolved, we will be able to talk about weapons that can bring victory in the war and create the conditions for the penetration of larger army groups on the front line.

Future tests will require the test center to be equipped with equipment that will be able to respond to the test tasks of characteristic drones driven by artificial intelligence and not one but rows. Basic tests will remain an important segment of the overall test but the control systems and ways of intercepting communications between operators and drones or their communications with AI will become a matter of survival on the front.

References

[1] https://orbotix.tech/autonomous-defense-2026/

[2] https://benchmark.rs/vesti/nauka-i-tehnologija/kina-prikazala-kompletno-funkcionalan-atlas-sistem-koji-za-operacije-koristi-roj-dronova/.

[3] https://support.turtlebeach.com/s/article/Atlas-Air-Swarm-II

[4] https://www.war.gov/News/News-Stories/Article/Article/4495825/defeating-the-swarm-project-flytrap-accelerates-natos-counter-drone-lethality

[5] https://publications.sto.nato.int/publications/STO%20Technical%20Reports/RTO-AG-300-V17/AG-300_V17-$$ALL.pdf

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CRITICAL LOAD CALCULATION FOR SOME KIND OF NONUNIFORM STRUTS

BOGDAN S. BOGDANOVIĆ
SERBIA, bogdanovic00@gmail.com

TONKO A. MIHOVILOVIĆ
UTVA AI, SERBIA, tonkojetonko@gmail.com

Abstract: *Struts of different types and shapes, made of different materials, are very freequently used as structural elements in aircraft and some other constructions. They are exposed to different types of loadings, and their dimensions should defined so (to be capable) to have sufficient capacity for such load transfering. Perhaps, the most important detail (data) for dimensions determination is definition of critical load (force). In this paper one of the well known numerical method for critical load definition, double integrated method, will be applied to some special kind of struts, having nonuniform cross section shapes This kind of struts isn´t often used, but it appears sometimes in constructions, and usualy it was solved using some kind of rough approximations. Double integrated method is quick and more reliable procedure for critical load (force) calculation. Also a numerical examples are given.*

Key words: *critical load, strut, nonuniform shape*

1. INTRODUCTION

In the greatest numbers of load cases, structural elements like struts are loaded in axial direction; these loads produce tension or compression in element. In the case of axial tension load transfering dimensions definition is quite simple. Here it depends only of strut´s cross section shape and dimensions; length of strut doesn´t have any influence. On the other side, in the case of axial compression load transfering, the length of strut must also be taken in consideration. Length of strut together with a cross section dimensions will be used for definitin of critical load (force), or BUCKLING load value; which is always the most important in considering struts load transfering capacity. Usually, struts have uniform cross section area shape along the whole length, but in some special cases, it could be partly changed; such struts type is named nonuniform. For both struts types uniform and nonuniform the double integration method could be used for BUCKLING load value definition.

2. DOUBLE INTEGRATION METHOD

This method could be used quite easy for critical load (force), or BUCKLING load calculation (estimation) of uniform, or nonuniform pin ended struts, but it loses simplicity for non pin ended struts. It should be noted, that pin ended boundary conditions always give the lowest critical load (force), or BUCKLING load value, so the obtained results could be used as a minimum value. It means, obtained results are conservative, on the safety side. The method ivolves a double integration of an assumed deflection curve. In [1] the differential equation for the displacement (w) of pin ended strut with a nonuniform cross section(flexuralrigidity)is:

$$Eiw'' + Pw = 0$$

Hence

$$w'' = -\frac{P*w}{EI} \qquad (1)$$

Assumed deflected shape (w_1) is a first step in calculation. Here P is treated as unknown quantity. It means that equation (1) could be used to determine the curvature w´´ in terms of P at any point x. Curvature is integrated twice in order to to obtain a new set (shape) of deflection, (w_2) which include P as a common factor. Dividing assumed deflected shape (w_1) with a new deflection set (w_2) the estimated value of P is obtained. This P value can maintain the assumed deflection (w_2); it means it is the critical load (force). There are two types of nonuniform shape struts:

- Cross section (EI) varies in steps, one or more, along the strut length (variant a), and (or)
- Cross section (EI) varies continously along the strut length (variant b)

Corresponding author: B. Bogdanović;
E-mail: bogdanovic00@gmail.com
DOI: 10.5937/OTEH26011B

These two types are quite similar, but the method procedure will be shown using strut where cross section varies in steps, because there it is less complicated and easier for understanding. For simplicity whole calculation is divided in several steps:

1. The strut is divided into several (usualy ten) segments of equal length), but because of symetry, only one half of strut´s length will be used,
2. Ends of segments are doubled in places where are stepped cross sectional changes
3. Assumed deflection curves are sine or parabola; maximum deflection (amplitude) is labeled as 100; it´s position is in the middle point of strut´s length.
4. Values of relative cross section characteristics EI are specified, at the ends of segments,
5. (w/EI) value calculation at the ends of strut segments,
6. Area under (w/EI) line calculation (using trapezoidal rule for every segment,
7. Slope change calculation along each segment $i_x = \int_0^x \left(\frac{P*w}{E*I}\right) dx,$
8. Obtaining (w´) cumulative values of (i_x) on the segments ends
9. Slope values (displacement changes w´) calculation at the segments ends
10. Area under (w´) line calculation (using trapezoidal rule for every segment)
11. Obtaining displacement values (w2) simply by taking cumulative total of numbers in preeceding row (step),
12. Finaly, an estimated value for relative critical load (force) in every segment can be obtained by dividing displacement values of assumed deflection curve (w1) with modified values (w2) and multiplying this ratio by 100

Results obtained in 12. th step will be different from segment to segment, and should be treated as a first iteration. Good aproximation for relative strut critical load (force) final result could be the average values (w1/w2) for all ten segments. Perhaps, it is better to take one or several more calculation cycles, (second, third iteration), taking modified value (w2) from preceeding iteration as assumed value (w1) for the next one:

$$w2_{n} = (w1_{n+1})*w_{\text{STRUT middle place}}$$

The procedure should be stopped when obtained results become the same for all segments. Results obtained in 12 th row of last iteration are also relative values, and to get absolute (real) strut critical load (force) it is necessary to multiple them with (EI/h2); it is obtained using equation:

$$P = (w1/w2)*(E*I)/h^2 \qquad (2)$$

Replacing h = L/10 in equation (2), it is obtained:

$$P = 100 * (w1/w2)*(E*I)/L^2 \qquad (3)$$

Here is: L = length of whole strut.

Now, it should be reminded that critical force for uniform cross section strut is defined by [2]:

$$P_{UNIFORM} = \pi^2*E*I/L^2 \qquad (4)$$

Comparing values obtained using equation (3) and (4), a reduction factor (coefficient) for strut critical load force produced by cross section nonuniformity is obtained:

$$\mu = (100 *(w_1/w_2))/\pi^2 \qquad (5)$$

3. NUMERICAL EXAMPLE

As a numerical example a critical load (force) for two types of nonuniform struts will be calculated. They both have equal length, and are made of the same material - alluminium alloy:

L = 100 [cm] , E = 749000 [daN/cm^2]

One of them is:

- reinforced in the middle section, STRUT 1, and the other is
- reinforced near the ends (supports) STRUT 2

Changes of their shapes along length are shown schematicaly in Fig. 1 and 2.

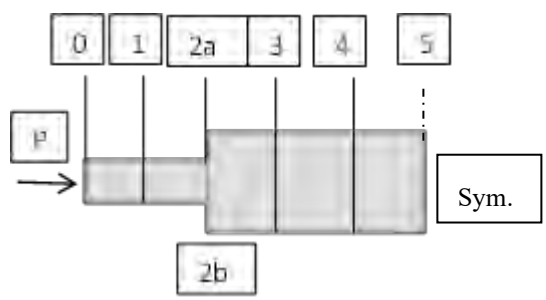


Figure 1 STRUT 1 - Shematic diagram

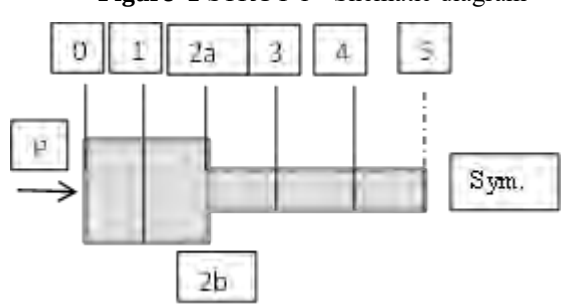


Figure 2 STRUT 2 - Shematic diagram

3.1 STRUT 1 critical force calculation

Calculation procedure and obtained results for STRUT 1 is shown in tables 1 - 4, 1 - 4 (cont), All necessary relative dimensions for STRUT 1 are shown in the second row in tables 1 - 4, and tables 1 - 4 continued

Table 1 First iteration of STRUT 1 critical load calculation

Pos.	0	0-1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.3142		0.628	0.6283
w1 [%]	0		0.3090		0.588	0.5878
EI	0.4		0.4		0.4	1
w/EI	0		77.2542		146.95	58.78
area		38.63		112.10		
ix	0		38.627		150.73	150.73
w'	406.12		367.497		255.40	255.40
area		386.81		311.45		
w2	0		386.810		698.26	698.26
P	0		0.0799		0.0842	0.0842

Table 1 cont First iteration of STRUT 1 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.94		1.26		1.57
w1 [%]		0.81		0.95		1
EI		1		1		1
w/EI		80.90		95.11		100
area	69.84	0	88.00	0	97.55	0
ix		220.57		308.57		406.12
w'		185.56		97.55		0
area	220.48		141.56		48.78	
w2		918.73		1060.29		1109.06
P		0.0881		0.0897		0.0902

Obtained results for P are sgnificantly different from segment to segment so the calculation will be continued thruogh Tables 2 and 2 (cont) – second iteration

Table 2 Second iteration of STRUT 1 critical load calculation

Pos.	0	0-1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.3142		0.628	0.628
w1 [%]	0	0	0.3488	0	0.630	0.630
EI	0.4		0.4		0.4	1
w/EI	0		87.19		157.40	62.96
area		43.60		122.30		
ix	0		43.60		165.89	165.89
w'	425.81		382.22		259.92	259.92
area		404.01		321.07		
w2	0		404.01		725.08	725.08
P	0		0.0863		0.0868	0.0868

Table 2 cont Second iteration of STRUT 1 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.94		1.26		1.57
w1 [%]	0	0.83	0	0.96	0	1
EI		1		1		1
w/EI		82.84		95.60		100
area	72.90	0	89.22	0	97.8	0
ix		238.79		328.01		425.81
w'		187.02		97.80		0
area	223.5		142.4		48.9	
w 2		948.55		1090.96		1139.86
P		0.0873		0.0876		0.0877

Obtained results for P are also a little bit different from segment to segment so the calculation will be continued through Tables 3 and 3 (cont) - third iteration.

Table 3 Third iteration of STRUT 1 critical load calculation

Pos.	0	0-1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.3142		0.628	0.628
w1[%]	0	0	0.3643	0	0.654	0.654
EI	0.4		0.4		0.4	1
w/EI	0		91.08		163.44	65.38
area		45.5		127.26		
ix	0		45.55		172.79	172.79
w'	440.76		395.23		267.97	267.97
area		418		331.60		
w2	0		418.00		749.60	749.60
P	0		0.0871		0.0872	0.0872

Table 3 cont Third iteration of STRUT 1 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.94		1.26		1.57
w1 [%]	0	0.86	0	0.98	0	1.03
EI		1		1		1
w/EI		85.53		98.37		102.78
area	75.45	0	91.95	0	100.6	0
ix		248.25		340.19		440.77
w'		192.52		100.57		0
area	230.3		146.6		50.29	
w 2		979.84		1126.39		1176.68
P		0.0873		0.0873		0.0873

For obtaining absolutely equal result for all segments it is necessary to make one more (fourth) iteration

Table 4 Fourth iteration of STRUT 1 critical load calculation

Pos.	0	0-1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.3142		0.628	0.628
w1 [%]	0	0	0.3769	0	0.676	0.676
EI	0.4		0.4		0.4	1
w/EI	0		94.22		168.97	67.59
area		47.11		131.6		
ix	0		47.11		178.71	178.71
w'	455.5		408.35		276.75	276.75
area		431.9		342.6		
w 2	0		431.91		774.46	774.46
P	0		0.0873		0.0873	0.0873

Table 4 cont Fourth iteration of STRUT 1 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.94		1.26		1.57
w 1 [%]	0	0.88	0	1.02	0	1.06
EI		1		1		1
w/EI		88.35		101.56		106.10
area	77.97	0	94.96	0	103.83	0
ix		256.68		351.63		455.46
w'		198.78		103.83		0
area	237.8		151.31		51.91	
w 2		1012.23		1163.53		1215.45
P		0.0873		0.0873		0.0873

Struts cross section shape is square, and maximum cross section dimensions, area and inertia are shown in Fig. 3, and table 8.

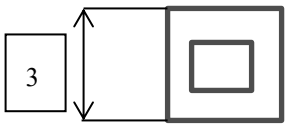


Figure 3 STRUTS 1,2 – maximum 100%, and reduced 40% strut cross sections

3.2. STRUT 2 critical force calculation

Calculation procedure and obtained results for STRUT 2 is shown in tables 5,6 and 5,6 (cont). The procedure

is absolute the same as for STRUT 1. All necessary relative dimensions for STRUT 2 are shown in the second row in tables 5, 6, and table 5, 6 continued. For STRUT 2, only the results from the first and the last (fourth) iteration will be shown.

Table 5 First iteration of STRUT 2 critical load calculation

Pos.	0	0- 1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.314		0.628	0.628
w1 [%]	0		0.309		0.588	0.588
EI	1		1		1	0.4
w/EI	0		30.90		58.779	146.95
area		15.5		44.84		
ix	0		15.45		60.291	60.29
w'	699.		683.3		638.49	638.49
area		691		660.9		
w 2	0		6911		1351.7	1351.7
P	0		0.0447		0.0435	0.0435

Table 5 cont First iteration of STRUT 2 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.942		1.26		1.57
w1 [%]		0.809		0.95		1
EI		0.4		0.4		0.4
w/EI		202.25		237.76		250
area	175.	0	220.0	0	243.9	0
ix		234.89		454.90		698.78
w'		463.89		243.88		0
area	551		353.9		121.9	
w 2		190376		2257.1		2379.0
P		0.0425		0.0421		0.0420

Table 6 Last (fourth) iteration of STRUT 2 critical load calculation

Pos.	0	0- 1	1	1--2	2a	2b
x [cm]	0		10		20	20
x [%]	0		0.1		0.2	0.2
α [rad]	0		0.3142		0.62	0.628
w 1 [%]	0	0	0.2848	0	0.58	0.558
EI	1		1		1	0.4
w/EI	0		28.476		55.76	139.39
area		14.24		42.12		
ix	0		14.238		56.35	56.35
w'	678.8		664.49		622.37	622.37
area		671.6		643.4		
w 2	0		671.61		1315.0	1315.0
P	0		0.0424		0.0424	0.0424

Table 6 cont Last (fourth) iteration of STRUT 2 critical load calculation

Pos.	2--3	3	3--4	4	4--5	5
x [cm]		30		40		50
x [%]		0.3		0.4		0.5
α [rad]		0.942		1.26		1.57
w 1 [%]	0	0.786	0	0.933	0	0.984
EI		0.4		0.4		0.4
w/EI		196.46		233.24		245.94
area	167.98	0	214.85	0	239.59	0
ix		224.28		439.13		678.73
w'		454.45		239.59		0
area	538.41		347.02		119.80	
w 2		1853.45		2200.46		2320.26
P		0.0424		0.0424		0.0424

Table 8 Dimensions of full strut cross section

b [cm]	h[cm]	h^3[cm]	I [cm^4]
3	3	27	6.75

Using equations (3), (4) and values from Tables 4, 5 and 7 the real STRUT 1 and STRUT 2 (nonuniform cross

section) critical force and critical force for uniform cross section strut are :

P = 100*0.0873*749000*6.75/100^2, Pcrit. = 4413.67 [daN] (nonuniform cross section) STRUT 1

P = 100*0.0424*749000*6.75/100^2,Pcrit. = 2122,417 [daN] (nonuniform cross section) STRUT 2

P = 3.14^2*749000*6.75/100^2, P = 4989.83 [daN] (uniform cross section)

4. CONCLUSION

The simplified and reliable procedure for critical force (load) values definition for struts having non uniform cross section shapes along length is shown. Only limiting condition for applying of this procedure is that struts are always pin ended. It could be applied to any kind of non uniformity, it means that cross section shapes can vary in steps, perhaps in several steps, or continously. Here the integration procedure is realized using trapezoidal rules; perhaps something better accuracy could be achieved using some other procedure (fitting cubic curve to the deflection line through segments). Two numerical examples are shown in detail.

References

[1] H.G.Allen, P.S. Bulson *Background of Buckling*, McGraw – Hill Book Company 1980

[2] E.F. Bruhn *Analysis and Design of Flight Vehicle Structures* Tri – State Offset Company Cincinnati Ohio 45202

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PLATE SPRING-BASED NEGATIVE STIFFNESS ROTARY MECHANISM FOR PASSIVE ENERGY BALANCING OF MORPHING WINGS

MAŠAN VUČINIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, 1088-2023@studenti.mas.bg.ac.rs

TONI IVANOV
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, tivanov@mas.bg.ac.rs

JELENA SVORCAN
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, jsvorcan@mas.bg.ac.rs

Abstract: *Increased actuation energy of morphing wings, caused by elastic deformation, could potentially diminish the aerodynamic benefits they offer compared to wings with conventional control surfaces. This paper proposes a rotary mechanism based on pre-buckled plate springs that exhibits negative stiffness and stores energy to balance the torque required for camber morphing. Plate springs benefit from near-constant post-buckling force output and could potentially simplify the negative stiffness mechanism. A combined analytical and numerical method for sizing the mechanism for camber morphing wings is developed and verified, accounting for stiffness requirements, range of motion, material properties, spatial constraints in small UAVs, and spring stability. Results indicate that the proposed mechanism is a viable solution for passive energy balancing of camber morphing wings.*

Keywords: *negative stiffness, morphing wing, energy balancing, compliant.*

1. INTRODUCTION

One of the challenges associated with morphing wings is the increased actuation energy required. This is due to the need to overcome the elastic forces caused by the deformation of the morphing wing structure. Consequently, larger actuators are required, increasing the overall system weight and potentially diminishing the aerodynamic performance benefits enabled by the smoother morphing geometry [1].

Morphing wing mechanism designs and concepts differ greatly in the materials used, types of morphing, and complexity. Camber morphing wings actuated by standard servo motors are of particular interest in this paper due to their simplicity and use of conventional means of actuation. Many variations exist, such as the FishBAC mechanism inspired by a fishbone [2], the corrugated spine mechanism [3], or the mechanism based on additively manufactured compliant composites, which was demonstrated in UAV flight [4].

Multiple solutions have been proposed to counteract the elastic deformation forces of morphing wings, specifically those actuated through a pulley and tendon mechanism. One such solution was a negative stiffness torsional mechanism proposed by Zhang et al. [5], which utilized pre-compressed springs arranged in a rotary mechanism. Wang et al. [6] opted for a somewhat more complex mechanism using a custom spiral pulley and extension springs.

1.1. Proposed Concept

This paper presents another solution for static balancing of an elastic camber morphing wing. It is based on the principle of strain energy conservation through pre-buckled plate springs. The springs are connected to a rotating shaft, as shown in Figure 1, providing the negative stiffness needed to counter the positive stiffness of the deforming wing, as the wing's deformation also produces torque on the shaft.

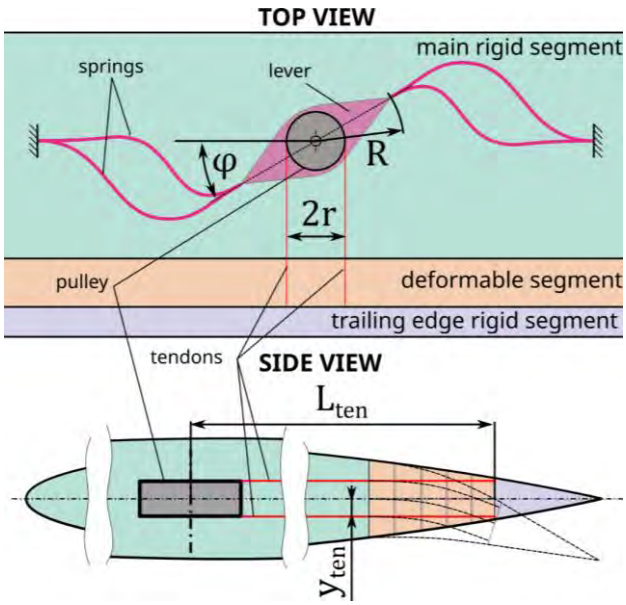


Figure 1. Proposed plate spring rotary mechanism morphing wing implementation

Corresponding author: M. Vučinić;
E-mail: 1088-2023@studenti.mas.bg.ac.rs
DOI: 10.5937/OTEH26012V

The lever and the pulley are tied and rotate together. In an ideal case, positive elastic deformation torque, acting on the mechanism through the pulley, and negative stiffness mechanism torque would cancel each other. This effect can be illustrated by the linear mechanism shown in Figure 2.

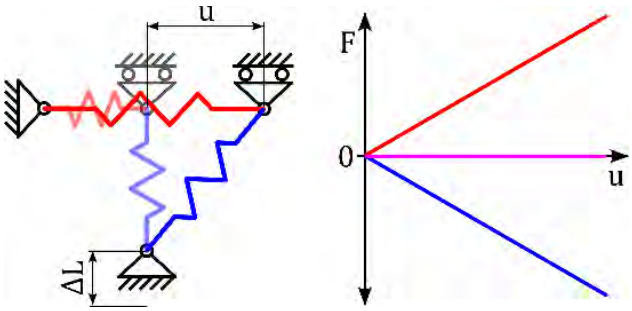


Figure 2. Negative stiffness mechanism based on a pre-compressed spring (blue) coupled with positive stiffness (red) yielding zero stiffness (magenta)

One of the goals of the proposed mechanism is to improve the scalability and simplicity in comparison to the previously mentioned works, as it would not depend on off-the-shelf components such as helical springs and rotary joints, and could even be manufactured as a monolithic piece.

A mixed analytical–numerical method is developed for designing such a mechanism and is demonstrated on a case study wing.

2. ANALYTICAL METHOD

2.1. Reference Wing Properties

Even though the proposed mechanism is not limited to a specific camber morphing design, for the purpose of this study, the FishBAC mechanism was chosen for the morphing wing segment due to its simplicity and wide literature coverage. The analytical method proposed by Woods et al. [7] was used to obtain a rough estimate of the stiffness and spatial constraints of the FishBAC mechanism, for which the negative stiffness mechanism was developed. This method is based on the reduction of the FishBAC model to a simple Bernoulli cantilever beam with variable stiffness along its length and calculation of the resulting deflection and needed torque for the given tendon material and cross-sectional area, where tendons are modeled as axially loaded rods. It is important to note that the required torque and pulley rotation for the given deflection depend strongly on the tendon material and dimensions:

$$M_{\text{ten}} = \frac{2k_{\text{ten}}y_{\text{ten}}\delta_R}{1 + 2k_{\text{ten}}y_{\text{ten}}\int_0^{L_m}\frac{1}{E(x)I(x)}\,dx} \tag{1}$$

where δ_R represents the change of tendon length due to the pulley rotation only, $E(x)$ and $I(x)$are the distribution of the elastic modulus and the axial moment of inertia of the equivalent beam along its length, and

$$k_{\text{ten}} = (E_{\text{ten}} * A_{\text{ten}}) \,/\, L_{ten} \tag{2}$$

represents tendon stiffness. The remaining terms from equations 1 and 2 are shown in Table 1.

Results were verified against the source (Woods et al. [7]) using the original input data, but for further analysis most of the dimensions and material properties were taken from the more recent work of Wu and Woods [8] and are presented in Table 1. The authors analyzed the wing at a freestream velocity of 30 m/s and demonstrated the benefits of a shorter morphing region near the trailing edge. As the method presented by Woods et al. [7] works only for symmetrical airfoils, the NACA 23012 airfoil from the work of Wu and Woods [8] was replaced with NACA 0012. Figure 3 shows the layout of the analyzed morphing segment and its maximum deflection.

The resultant moment-pulley angle relationship is linear, as expected. The maximum torque needed for the given parameters is $M_{\text{ten}} = 390$ N mm, occurring at the pulley angle $\varphi = 15°$, and the maximum deflection of the morphing segment end of $u_{\text{max}} = 3.3$ mm. Note that this deflection does not represent the trailing edge deflection as the morphing section stops at 90% of the chord; therefore, the actual wingtip deflection is larger. Additionally, it was estimated that the springs should not be wider than 25 mm due to spatial constraints.

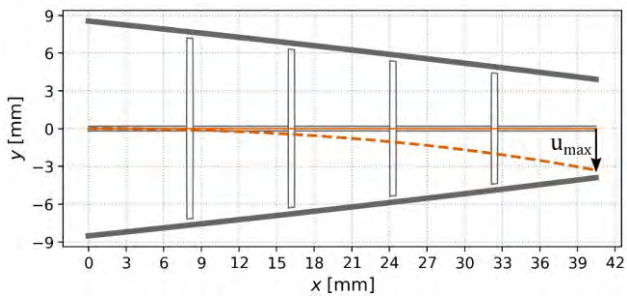


Figure 3. Morphing segment dimensions and calculated deflection for the pulley angle of 15°

Table 1. Reference wing parameters

Name	Symbol	Value
chord	c	270 mm
section span	b	1 m
morph start	x_s	0.75 c
morph end	x_e	0.9 c
pulley radius	r	3 mm
tendon length	L_{ten}	170 mm
stringer count	n	4
tendon offset from spine	y_{ten}	3 mm
tendon section diameter	D_{ten}	0.7 mm
spine thickness	t_{sp}	0.375 mm
stringer thickness	t_{st}	0.5 mm
skin thickness	t_{sk}	0.5 mm
spine elastic modulus	E_{sp}	8.58 GPa
stringer elastic modulus	E_{st}	2.01 GPa
skin elastic modulus	E_{sk}	1.22 MPa
tendon elastic modulus	E_{ten}	131 GPa
required torque	$\mathbf{M_{ten}}$	**390 N mm**

2.2. Spring Analysis

Based on the required torque and the maximum pulley rotation angle calculated in the previous chapter, it is now possible to size the negative stiffness mechanism. The rotational mechanism presented in Figure 4 consists of two springs approximated as two fixed-guided beams. The actual mechanism would have springs on both sides of the shaft, making it axisymmetric and canceling out the transverse forces acting on the shaft. Therefore, the model can be reduced to a single side and designed for half of the required torque, since the full system is symmetric and would double the torque output.

Dijksman [9] presented a detailed simplified linear analytical method for the analysis of translational mechanisms utilizing pre-buckled plate springs of undeformed length L and stiffness $B = (Ewt^3)/12$, where E is the elastic modulus, w is the spring width, and t is the spring thickness. According to their work, the moment M, force D parallel to the translation direction v in Figure 4, and force N perpendicular to the translation direction, all acting at the moving end, can be calculated as

$$D_i = \alpha_i^2 \left(\frac{\theta \sin\left(\frac{1}{2}\alpha_i\right) + \alpha \frac{v}{L} \cos\left(\frac{1}{2}\alpha_i\right)}{2\left(\sin\left(\frac{1}{2}\alpha_i\right) - \frac{\alpha_i}{2}\cos\left(\frac{1}{2}\alpha_i\right)\right)} \right) \cdot \frac{B}{L^2}, \quad (3)$$

$$N_i = \frac{B\alpha_i^2}{L^2}, \quad (4)$$

$$\frac{M_i L}{B} = \alpha_i \frac{\left(\theta\left(\frac{\frac{\alpha_i}{2}}{\sin\left(\frac{\alpha_i}{2}\right)} - \alpha_i \sin\left(\frac{\alpha_i}{2}\right) - \cos\left(\frac{\alpha_i}{2}\right)\right) - \alpha_i \frac{v}{L} \sin\left(\frac{\alpha_i}{2}\right)\right)}{2\left(\sin\left(\frac{\alpha_i}{2}\right) - \frac{\alpha_i}{2}\cos\left(\frac{\alpha_i}{2}\right)\right)}. \quad (5)$$

Here $i = 1,2..$ represents the spring index, θ is a constant initial spring angle at the sliding end of the mechanism, and $\alpha_i = 2(\pi - \varepsilon_i)$, where

$$\varepsilon_i = \pm \frac{1}{4}\theta \sqrt{\frac{L}{L - L_{0t}}} \quad (6)$$

depends on how bent the spring is at a given moment. In Figure 4, the left spring ($i = 1$) is more bent at $v = 0$ with given θ than in the case where $\theta = 0$, therefore $\varepsilon_1 < 0$. The right spring ($i = 2$) is less bent than in the $\theta = 0, v = 0$ case, therefore $\varepsilon_2 > 0$.

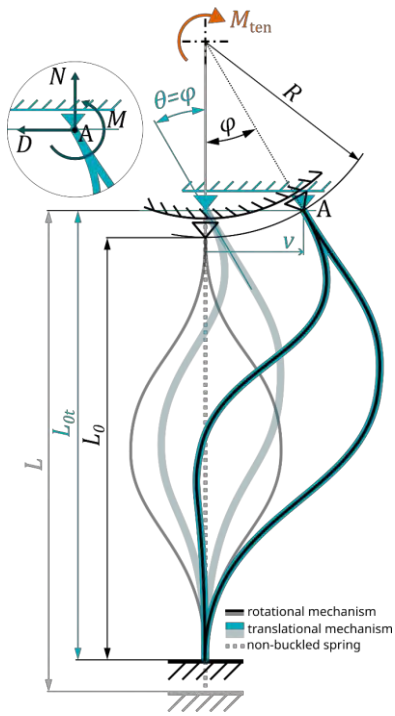


Figure 4. The rotating mechanism with its equivalent translational mechanism at the given moment

Given its path independence, the above-mentioned approach is modified here to account for rotational motion instead. The forces and moments of the rotational mechanism analyzed here and shown in Figure 4, with lever arm radius R and lever angle φ, can be calculated by substituting

$$v = R \sin \varphi, \quad (7)$$

$$L_{0t} = L_0 + R(1 - \cos \varphi), \quad (8)$$

$$\theta = \varphi, \quad (9)$$

into equations 3, 4, and 5.

Therefore, it is possible to calculate the moment generated in the pulley shaft by each spring of the negative stiffness mechanism as follows:

$$M_{Ri} = -\left(M_i + R(\sin \varphi \, N_i - \cos \varphi \, D_i)\right) \quad (10)$$

The minus sign is due to the tendon moment's positive direction defined in Figure 4. The more bent spring is inherently unstable. The left spring in Figure 4, in both the translational and rotational mechanism, would lose its stability after a certain displacement or rotation. It would snap to the same shape as the less bent spring on the right. Dijksman [9] also presented an equation

$$v_{\max,\mathrm{mb}} = 2L \sqrt{\frac{L - L_{0t}}{3L}} \left(1 - \frac{L - L_{0t}}{L}\right) \left(1 - 0.6\varphi \sqrt{\frac{L}{L - L_{0t}}}\right) \quad (11)$$

where $v_{\max,\mathrm{mb}}$ represents the maximum displacement v that the observed translational system can have before the more bent spring becomes unstable. The author also presented the equation to calculate the maximum displacement at which the less bent spring loses its spring action:

$$v_{\max,\mathrm{lb}} = 2L\sqrt{\frac{L-L_{0t}}{3L}}\left(1-\frac{L-L_{0t}}{L}\right)\left(1-0.26\varphi\sqrt{\frac{L}{L-L_{0t}}}\right). \quad (12)$$

It can be noted that in the case of the rotational mechanism only the more bent spring ($i = 1$) is the limiting factor, as $\varphi = \theta$ and v always have equal signs.

3. NUMERICAL ANALYSIS

3.1. Geometry Optimization

As the bending stress of a buckled spring is proportional to the change in length of the spring after buckling, it is essential to set adequate limits during the design phase so that the springs are stable but not unnecessarily overstressed.

In the mathematically ideal case, the more bent spring should be at its stability limit for the maximum pulley angle. This is impractical and, in aerospace applications, even unsafe, as the mechanism is very sensitive to any input around the stability boundary. For this reason, a safety factor of around 1.1 was implemented to ensure that the mechanism does not approach the stability boundary during operation.

The factor that limits the working range of the rotary mechanism is $v_{\max,1}$. Therefore, the following method was developed to find the combinations of relative changes in length and relative radii for which the mechanism operates at the stability limit for the given lever angle φ.

The contours shown in Figure 5 were obtained by computing the level set

$$q_1\left(\frac{R}{L}, \frac{\Delta L}{L}, \varphi\right) = 1 \quad (13)$$

where q_1 depends on the properties of the equivalent translational mechanism for the given R/L, $\Delta L/L$, and φ in a way that

$$q_1 = \frac{v_{\max,1}}{v} \quad (14)$$

The results are shown in Figure 5. The maximum pulley angle φ the spring can sustain depends solely on its length and the buckling displacement. Several points along the $\varphi \approx 16.6°$, equal to around 110% of the needed 15° of working range, were selected for the FEM analysis described in the following section, with the goal of determining output torque and stresses for each combination.

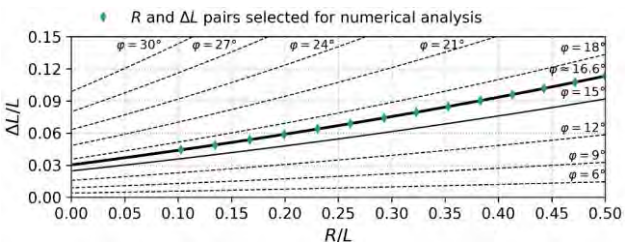


Figure 5. Curves of radius-length combinations at loss of stability for constant lever angle

3.2. FEM Analysis

An iterative static structural analysis was performed using Abaqus solver for each combination of R/L, $\Delta L/L$ and L, where L is a set of undeformed spring lengths shown in Table 2. Spring width and thickness were held constant at each iteration, as shown in the same table.

Table 2. Spring dimensions

spring width	w	25 mm
spring thickness	t	0.2 mm
spring undeformed length	L	[110, 120, 130, 140] mm

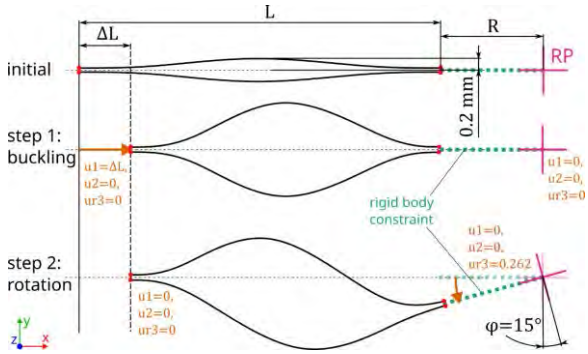


Figure 6. FEM analysis steps

Each analysis consisted of three steps shown in Figure 6. The beams were of a slightly sinusoidal shape with an amplitude of 0.2 mm (as shown in the initial step in Figure 6), to induce buckling in the right direction. Left end nodes of the springs had a displacement boundary condition prescribing translation to induce buckling in the first step (step 1 in Figure 6), and the remaining degrees of freedom constrained to zero. A rigid body constraint was used to connect the right end nodes to a reference point (RP) positioned to represent the pulley axis. The RP had a displacement boundary condition prescribing rotation (0° during buckling, 15° in the following step, as shown in Figure 6), and the remaining degrees of freedom were constrained to zero.

The model was two-dimensional. Linear Timoshenko beam elements in 2D space with the hybrid formulation (B21H) were used due to the simplicity of the geometry and computational efficiency. They were expected to have very similar results to those of Euler-Bernoulli theory, due to the very large slenderness of the beams. The hybrid formulation provided robustness during the initial buckling step. Mesh size was set to 2 mm. The large displacement solver option was used. The material used was linear elastic with the elastic modulus from Table 3.

Material properties of spring steel 1.4310, listed in Table 3, are based on typical values reported by spring steel manufacturers, used to assess the general viability of the proposed model rather than to represent a specific tested material batch.

Table 3. Assumed spring material properties

material name		1.4310 spring steel
Young's modulus	E	190 GPa
tensile strength	UTS	1600 MPa
fatigue strength	S_f	450 MPa

Reaction moment was calculated at the reference point, representing the shaft of the pulley. The rotation step maximum time increment was set to 0.05 to obtain the stress distribution and reaction moments at multiple points along the working range. The stress distribution for a selected configuration is shown on Figure 7 and Figure 8.

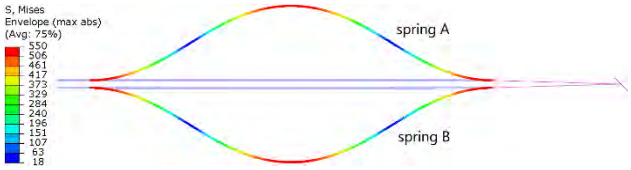


Figure 7. Maximum von Mieses stress (MPa) along the beams after the buckling step for $\Delta L = 8.9$ mm, $R = 35.14$ mm, $L = 120$ mm. The unbuckled spring and the rigid body are shown

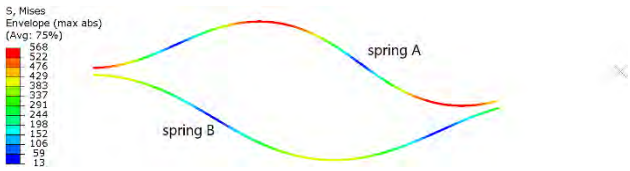


Figure 8. Maximum von Mieses stress (MPa) in the rotation step for $\varphi = 15°$, $\Delta L = 8.9$ mm, $R = 35.14$ mm, $L = 120$ mm

3.3. Mechanism Sizing

The FEM analysis results provide information about the torque exerted along the working range for each combination of R, L and ΔL from Figure 5 and Table 2. These results can be directly used to find the combinations that satisfy the torque requirement established in Section 2.1., or to be more precise, half of the torque required, as only one half of the mechanism is analyzed. However, raw stress values are not indicative of a given configuration's viability, as they do not consider fatigue life of the spring, which is a major design constraint.

To account for the fatigue effects, the Goodman criterion is selected as another condition to be satisfied together with the required torque. For the fatigue analysis, it is assumed that the mechanism cycles through its full range of motion each cycle. This is an intentional overestimation, as this would rarely happen during flight if the morphing structure were, for example, part of an aileron, and used as the theoretically worst possible case. In general form, the Goodman criterion is calculated based on the stress amplitude σ_a and the mean stress σ_m, as

$$\eta_\mathrm{G} = \frac{\sigma_\mathrm{a}}{\mathrm{S_f}} + \frac{\sigma_m}{\mathrm{UTS}} = \frac{\frac{\sigma_\mathrm{max} - \sigma_\mathrm{min}}{2}}{\mathrm{S_f}} + \frac{\frac{\sigma_\mathrm{max} + \sigma_\mathrm{min}}{2}}{\mathrm{UTS}} \quad (15)$$

where σ_max and σ_min represent maximum and minimum stress at the examined point.

It is important to note that the range of rotation in the FEM model is from 0° to 15°, which is only half of the actual working range from -15° to 15°. The mechanism's symmetry can be used to account for this. Beam B (spring B in Figure 8 and Figure 7) is a mirrored version of beam A (spring A in Figure 8 and Figure 7). In case $\varphi < 0$, beam A would take the mirrored shape of beam B for $|\varphi|$. This practically means that for $\varphi < 0$ beam A's top surface has values corresponding to beam B's bottom surface for $|\varphi|$. This enables the introduction of an equivalent spring for which the maximum Goodman criterion for each node j is equal to

$$\eta_{\mathrm{G}j} = \max\left(\frac{\sigma_{\mathrm{a}j}^{\mathrm{top}}}{\mathrm{S_f}} + \frac{\sigma_{\mathrm{m}j}^{\mathrm{top}}}{\mathrm{UTS}}, \frac{\sigma_{\mathrm{a}j}^{\mathrm{bot}}}{\mathrm{S_f}} + \frac{\sigma_{\mathrm{m}j}^{\mathrm{bot}}}{\mathrm{UTS}}\right) \quad (16)$$

The maximum overall Goodman criterion of the mechanism is now equal to

$$\eta_{G_\mathrm{max}} = \max_{j\in\{1,\dots,n\}} \eta_{G_j} \quad (17)$$

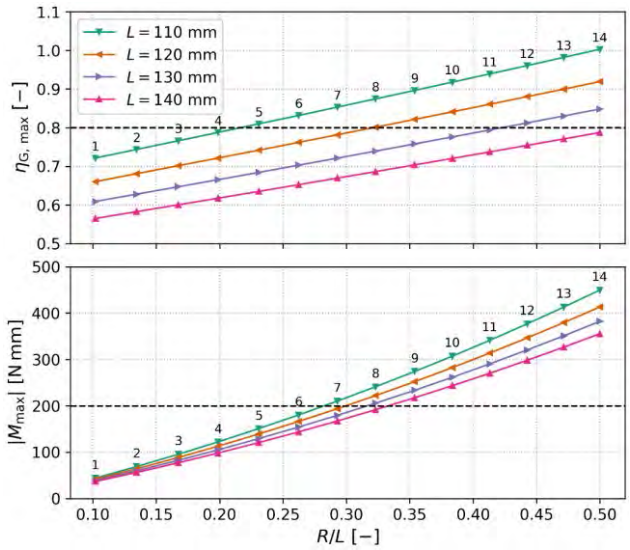


Figure 9. Maximum moment and Goodman criterion for each combination of R, L, and ΔL

The maximum moment measured at the reference point and the maximum Goodman criterion are shown in Figure 9 for each analyzed configuration.

Out of all the discretized combinations used for the FEM analysis, geometric combination 7 with $L = 120$ mm was selected for the experiment. This decision was based on the fact that it is the closest match to the required moment that doesn't overshoot it, with the moment of -193.6 Nmm compared to the 195 Nmm needed. Having the negative stiffness mechanism stiffer than the positive stiffness mechanism would produce overall negative stiffness which is undesired. The Goodman criterion value at just under 0.8, which is selected as the upper limit considering a factor of safety, also satisfies the fatigue condition.

4. EXPERIMENTAL VERIFICATION

The experimental setup, shown in Figure 10, consisted of an adjustable anchor used to fix one end of the springs after buckling them, an adjustable lever arm and a pulley segment, whose rotation was limited by the adjustable stopper, and a string attached to the pulley. The pulley radius was 25 mm and the string diameter was 1 mm.

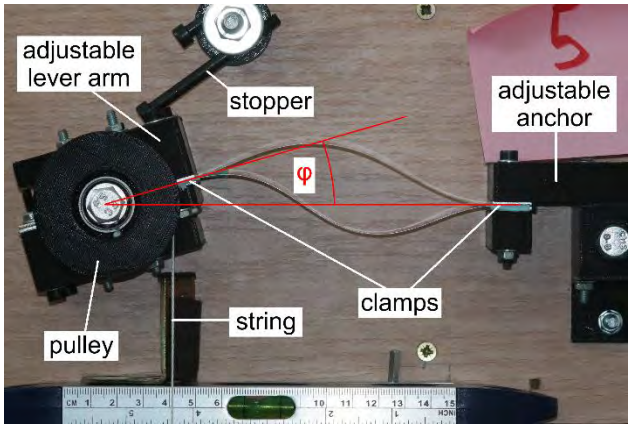


Figure 10. The experimental setup with its components

The springs were inserted in clamps and fastened to desired length, while keeping the pulley angle as close to zero as possible. The adjustable anchor was moved to buckle the springs. The stopper was adjusted to the desired position, and the mechanism was left to rest on it. Weight was incrementally added through the spring acting on the pulley until the system lost stability and snapped back to the other stable state.

The maximum moment at the given stopper position was calculated after measuring the weight with ± 3 g precision. The corresponding angle was measured using ImageJ from photographs taken with a telephoto lens with minimal distortion, as shown in Figure 10. These measurements were repeated for multiple stopper positions.

The springs used were cut from a sheet of 1.4310 steel, with the width of 25 ± 0.5 mm, and thickness of 0.2 ± 0.008 mm. The length of the springs between two clamps was set to $L \approx 120$ mm, the lever arm radius to $R \approx 35.1$ mm, and the translation of the adjustable anchor after clamping the springs was set to $\Delta L \approx 8.9$ mm. The results are shown in Figure 11.

5. RESULTS AND DISCUSSION

Results presented in Figure 11 show a strong match between numerical and analytical solutions for the moment curve, and somewhat lower values for the experimental curve. This difference is deemed acceptable and expected considering the measurement uncertainty mentioned in Section 4 and other effects that are not accounted for in the analysis, such as the stiffness of the experimental setup. Aside from that, the material batch used for the springs in the experiment was not tested, and slight variations in the Young's modulus are possible.

The relationship between the moment and the angle is nearly linear, and the measured stability boundary slightly above the calculated one, as expected due to the slight initial curvature of the springs used in the experiment.

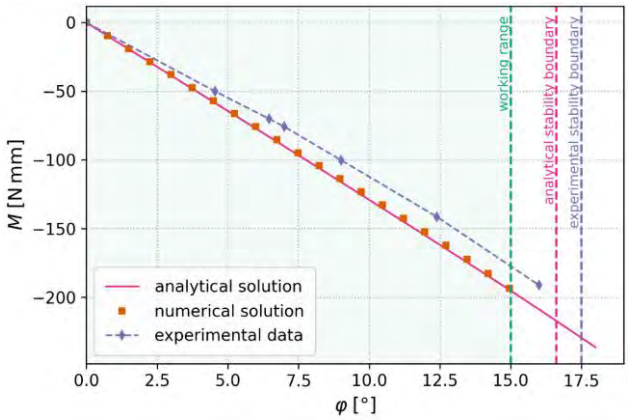


Figure 11. Calculated moment versus angle curves for $L = 120$ mm, $R = 35.14$ mm, $\Delta L = 8.9$ mm, $h = 0.2$ mm, $w = 25$ mm and $E = 190$ GPa, and measured experimental data for values stated in Section 4

It was observed that that the stiffness of the mechanism is highly sensitive to the change in initial spring length L. For example, increasing the spring length from 120 to 125 millimeters leads to strong agreement with the experimental results. At 16°, the model with $L = 125$ mm predicts a moment of -190.5 Nmm, compared to -191.1 Nmm from the experimental data, whereas the model with $L = 120$ mm predicts -208.9 Nmm. This effect can be used to tune the mechanism's stiffness during its assembly in the wing.

Nonetheless, the buckling displacement ΔL has a minor effect on the generated moment, but a major effect on the stability range of the spring and stress.

6. CONCLUSION

Overall, the constraints and requirements of the reference wing were successfully satisfied, proving the potential this mechanism has in morphing wing applications. The plate-spring mechanism decreased the net required actuation torque significantly, as shown in Figure 12.

The plate-spring based mechanism has lower part count and fewer connectors when compared to the helical spring mechanisms. Another opportunity plate-spring based system offers when compared to helical spring mechanisms is the possibility of additive manufacturing of a monolithic mechanism version.

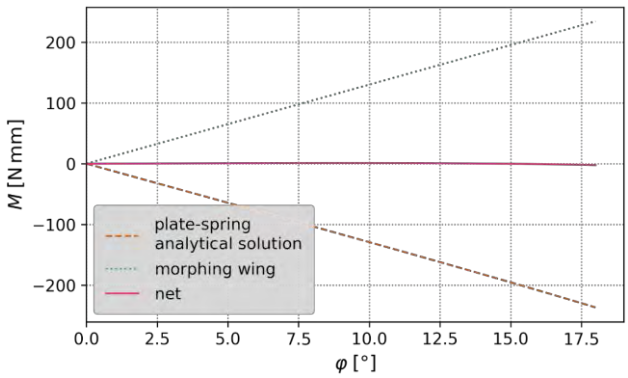


Figure 12. Required torque with the plate-spring mechanism

References

[1] R. Pecora, "Morphing wing flaps for large civil aircraft: Evolution of a smart technology across the Clean Sky program," *Chinese Journal of Aeronautics*, vol. 34, no. 7, pp. 13–28, Jul. 2021, doi: 10.1016/j.cja.2020.08.004.

[2] B. K. Woods, O. Bilgen, and M. I. Friswell, "Wind tunnel testing of the fish bone active camber morphing concept," *Journal of Intelligent Material Systems and Structures*, vol. 25, no. 7, pp. 772–785, May 2014, doi: 10.1177/1045389X14521700.

[3] T. Yokozeki, A. Sugiura, and Y. Hirano, "Development of Variable Camber Morphing Airfoil Using Corrugated Structure," *Journal of Aircraft*, vol. 51, no. 3, pp. 1023–1029, May 2014, doi: 10.2514/1.C032573.

[4] U. Fasel, D. Keidel, L. Baumann, G. Cavolina, M. Eichenhofer, and P. Ermanni, "Composite additive manufacturing of morphing aerospace structures," *Manufacturing Letters*, vol. 23, pp. 85–88, Jan. 2020, doi: 10.1016/j.mfglet.2019.12.004.

[5] J. Zhang, A. D. Shaw, M. Amoozgar, M. I. Friswell, and B. K. S. Woods, "Bidirectional Spiral Pulley Negative Stiffness Mechanism for Passive Energy Balancing," *Journal of Mechanisms and Robotics*, vol. 11, no. 5, p. 054502, Oct. 2019, doi: 10.1115/1.4043818.

[6] C. Wang *et al.*, "Integration of the passive energy balancing based actuation system into a camber morphing design," *Aerospace Science and Technology*, vol. 155, p. 109641, Dec. 2024, doi: 10.1016/j.ast.2024.109641.

[7] B. K. S. Woods, I. Dayyani, and M. I. Friswell, "Fluid/Structure-Interaction Analysis of the Fish-Bone-Active-Camber Morphing Concept," *Journal of Aircraft*, vol. 52, no. 1, pp. 307–319, Jan. 2015, doi: 10.2514/1.C032725.

[8] J. Wu and B. K. S. Woods, "Modelling and analysis of two-dimensional static and dynamic aeroelasticity of Fish Bone Active Camber morphing aerofoils," *Journal of Fluids and Structures*, vol. 129, p. 104171, Oct. 2024, doi: 10.1016/j.jfluidstructs.2024.104171.

[9] J. F. Dijksman, "A study of some aspects of the mechanical behaviour of cross-spring pivots and plate spring mechanisms with negative stiffness," PhD Thesis, TU Delft, 1979. [Online]. Available: https://resolver.tudelft.nl/uuid:12c3dceb-5aa9-4b07-96d6-1bde8e1c96a3

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DESIGN AND VALIDATION OF TOPOLOGY-OPTIMIZED 3D-PRINTED QUADCOPTER ARMS

MILAN Z. RAKIC
Innovation center, Faculty of Mechanical Engineering, Belgrade, mzrakic@mas.bg.ac.rs

TONI D. IVANOV
Faculty of Mechanical Engineering, Belgrade, tivanov@mas.bg.ac.rs

Abstract: *This paper presents an approach to topology optimization of a small quadcopter's structure, with particular emphasis on the optimization and load validation of the drone arm. The priority of this approach is to reduce the structural mass while maintaining adequate stiffness and load-bearing capacity, based on the propulsion requirements of the motor system. A finite element numerical model of the arm was developed and compared with an analytical calculation based on beam theory, in order to validate the simplified mathematical model. Topology optimization was carried out using the tools available in Ansys Workbench 25, reducing the arm's mass from 148 g to 106 g (~28%), while satisfying the imposed deflection and strength criteria. The optimized model has been prepared for manufacturing via FDM 3D printing, while fabrication and experimental validation will be the subject of future work.*

Keywords*: topology optimization, small quadcopter, 3D printing, optimization algorithm, structural integrity.*

1. INTRODUCTION

Bearing in mind that drones have no capability to remain airborne without propulsion (unlike gliders and aircraft, which possess an aerodynamic glide ratio), it is essential to address the issue of unnecessary drone mass. The largest share of a quadcopter's mass is taken up by the electric propulsion system [1]. Since we had a propulsion system that needed to be tested, we designed a small quadcopter capable of adequately accommodating all the components required for autonomous flight. Such a drone is expected to be compact, easy to manufacture, transport, and assemble. In order to remove the unnecessary mass of the drone structure that carries no load during flight, it was decided to subject the arm of the classic "X-configuration" to structural (topology) optimization.

The drone arm's role is to accommodate the electric motor with propeller and transfer its action to the frame. The greatest load on the arm is the bending moment caused by the motor thrust, which tends to bend it around the point where it connects to the structure. Due to the high rotational speed, the motor also transmits vibrations that, over prolonged operation, can cause fatigue of the arm material. When dimensioning the arm, the torsional moment about the longitudinal axis, caused by the eccentric thrust force during sharp turns, must also be taken into account.

The significance of topology optimization lies in its ability to produce a more economical, lighter, and cheaper product in virtually any branch of engineering. Its goal is to distribute material geometrically within a component so as to achieve a desired physical parameter under given conditions. In this paper, the structurally optimized model is required to have the lowest possible mass while retaining adequate stiffness. The first published scientific work on topology optimization was by the Danish researcher Bendsøe Martin Philip and co-workers [2].

The optimized geometry should be aesthetically more pleasing, simpler to use, or occupy less space than the original model. By applying the finite element method and the topology optimization theory explained in references [3, 4], software tools generate a model that satisfies the specified conditions. Such a model is often very demanding to manufacture using conventional processes (e.g., mold casting or machining) and requires further post-processing. In this case, it is most convenient to prepare the model for the FDM 3D printing process, which builds the part layer by layer.

Regarding the structure of a multirotor, the greatest potential for optimization lies in the arms and the frame. Removing excess, unnecessary mass saves battery energy and creates space for mounting the electronic components required for filming or flight control. For the selected electric propulsion system, it was necessary to design as light a drone as possible, with an arm span of about half a meter, intended to test the correct operation of the motors and to fly autonomously. The goal is to design an optimized arm with significantly reduced mass while retaining adequate stiffness and strength properties, so that the arm can withstand all loads caused by the operation of the propulsion system.

Corresponding author: M. Rakić;
E-mail: mzrakic@mas.bg.ac.rs
DOI: 10.5937/OTEH26013R

2. APPLICATION OF TOPOLOGY OPTIMIZATION TO THE DRONE ARM

2.1. Description of the Drone Structure and Arm Geometry

The drone is a quadcopter of the "X-configuration". Its propulsion system uses 4 electric motors coupled with electronic speed controllers (ESCs), powered by a 4-cell (4S) lithium-polymer battery. The motors are fitted with 10×4.5 propellers, whose diameter was one of the main factors in determining the arm spacing. The arms are bolted to an octagonal drone frame consisting of 2 plates: the top plate carries the Pixhawk PX4 autopilot and GPS module, while the upper side of the second plate holds the Power Management Board PM07 and the radio receiver, with the battery mounted on its underside.

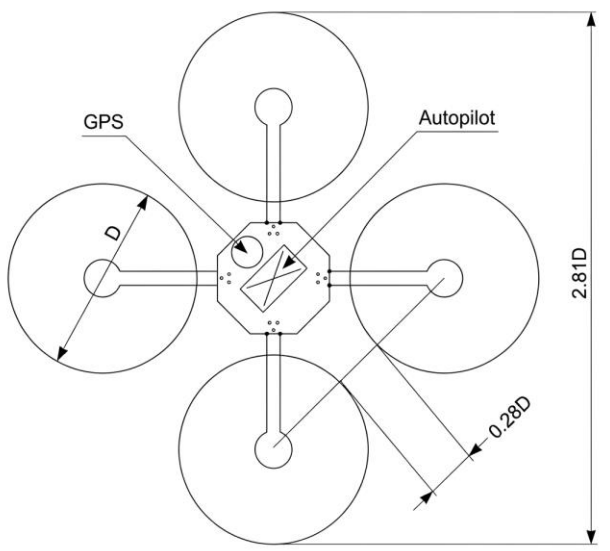


Figure 1. Drone sketch

When determining the arm span, it was considered that the distance between propellers should be as small as possible, so that the central body has room only for the arm mounts and the required components. On the other hand, this distance must not be too small, in order to avoid interaction between the propeller tips. Three M3 screws are used to clamp the arm to the frame. At the end of the arm is the flange for mounting the motor. The ESC controllers are positioned along the arms. The wiring of all electrical components has been shortened to simplify the connections as much as possible.

2.2. Analytical Model of the Arm

For dimensioning purposes, an analytical model of the arm is constructed. The arm is idealized as a cantilever beam clamped at the frame mount, loaded at its free end by a radial force representing the propeller thrust. A force value of $F = 3 \cdot g \approx 30$ N was adopted, corresponding to the manufacturer-specified value for the case in which the motor enters a safety (critical) operating regime. The point of application of this force is offset from the centroidal axis and positioned with an eccentricity of $e = 10.6$ mm from the center of the flange, at the location of the motor-mount hole. Such a position of the force application point can occur if the bolted connection between the mount and the flange loosens due to vibration or loading.

A square cross-section was chosen for the arm. The primary condition governing the dimensioning of the arm is the maximum-deflection condition, i.e., the maximum deformation of the free end must not exceed 1% of the arm length. The point of force application is located at a distance $l_F = 175$ mm from the arm's clamped end (the thrust acts at the center of the flange where the motor is mounted), giving an allowable deflection of $\delta_{perm} = 1.75$ mm. The second dimensioning condition is the allowable stress, comprising bending and torsional components (a combined stress state), due to the offset of the force with respect to the longitudinal axis of symmetry.

The arm is manufactured by the FDM (Fused Deposition Modeling) 3D printing process, using PLA-CF material. This material is a standard PLA plastic reinforced with carbon fibers, which improve the plastic's mechanical properties while retaining ease of manufacturing. The properties of the PLA-CF material, with a fiber length of 8 mm and a fiber content of 16% [5], adopted in the calculation are given in Tables 1-3. The material density is $\rho = 1760$ kg/m^3, and the tensile strength is $\sigma_T = 68$ MPa.

Table 1. PLA-CF material properties – Young's modulus

E_1 [MPa]	E_2 [MPa]	E_3 [MPa]
3774.8	2820.9	2976

Table 2. PLA-CF material properties – shear modulus

G_{12} [MPa]	G_{13} [MPa]	G_{23} [MPa]
1114	1121	983

Table 3. PLA-CF material properties – Poisson's ratio

ν_{12}	ν_{13}	ν_{23}
0.33	0.3	0.33

It is necessary to determine the unknown side length of the square cross-section of the arm based on the governing condition of allowable deflection at the free end. The deflection consists of a bending component and a shear component, and their sum must be less than the allowable deflection. It is calculated according to the equation:

$$\delta_{max} = \frac{F l_F^{\,3}}{\frac{3E_1 a^4}{12}} + \frac{F l_F}{k_s G_{12} a^2} \leq \delta_{perm} \qquad (1)$$

where $k_s = 0.833$ is the shear correction factor. A minimum required safety factor of $S_F = 1.1$ was imposed to account for imperfections inherent to 3D printing manufacturing. After carrying out the calculation taking these conditions into account, and rounding the side length up to the next value permitted by printing precision, the required side length is found to be $a = 18.2$ mm, with a deflection safety factor of $S_{def} = 1.115$. When checking the stress under the

combined loading state (bending and torsion), the cross-section is dimensioned according to the HMH (von Mises) criterion of maximum equivalent stress. The following four equations describe this process:

$$\sigma_{bM} = \frac{M_{max}}{W} = \frac{Fl}{\frac{a^3}{6}} \quad (2)$$

$$\tau_{tM} = \frac{T}{\alpha_t a^3} = \frac{Fe}{0.208 a^3} \quad (3)$$

where $\alpha_t = 0.208$ is the Saint-Venant torsion coefficient for a solid square cross-section.

$$\sigma_{ekv} = \sqrt{\sigma_{bM}^2 + 3\tau_{tM}^2} = 5.244 \text{ MPa} \quad (4)$$

$$S_{str} = \frac{\sigma_T}{\sigma_{ekv}} = 12.97 \quad (5)$$

This procedure yields a strength safety factor of $S_{str} = 12.97$, indicating that such a cross-section is over-dimensioned for this loading case. Since the maximum-deflection condition is governing, the side length is adopted according to this condition. With this calculation, the maximum deflection is $\delta_{max} = 1.57$ mm.

2.3. Theoretical Basis of Topology Optimization

The goal of topology optimization is to find, within a predefined design domain, boundary conditions, and loads, a material distribution that satisfies the imposed condition. The problem is formulated as a mathematical optimization task: an extremum of the objective function (e.g., compliance — a measure of structural flexibility) is sought, subject to the imposed constraints (mass/volume, displacement, etc.) [3].

From [2], the density-based method of structural optimization, known today as SIMP (Solid Isotropic Material with Penalization), evolved into the most widespread approach [3], [4]. Each finite element of the design domain is assigned a pseudo-density in the range from 0 to 1 (0 meaning no material, 1 meaning solid material), whereby the element stiffness is scaled by a power of this density using a penalization exponent.

An alternative approach is the Level-Set method, which, unlike the previous approach, does not operate on element-wise density, but instead directly describes the shape boundary as the zero contour (isosurface) of a scalar function defined over the whole domain [6]. By evolving this function through iterations, the shape boundary progresses toward the optimal solution. Also noteworthy is the Mixable Density method, which combines these two approaches — merging the smooth solution of the Level-Set approach with the flexibility of element-wise density.

Regardless of the chosen formulation, the resulting nonlinear constrained optimization problem is solved iteratively using appropriate numerical algorithms. For the classical case (compliance objective function, mass/volume constraint), the Optimality Criteria method is most commonly used, being well suited to this type of problem. When additional constraints are introduced — such as a local stress constraint — a more general solution approach is required, based on Sequential Convex Programming, within which several concrete algorithm implementations exist, such as the modified Method of Feasible Directions (MFD) and the SCPIP method, based on an interior-point approach [7].

2.4. Application of Topology Optimization in Ansys Workbench 25

The first step after launching Ansys Workbench 25 is to create a Static Structural analysis environment. The solution of this analysis must be linked to the Topology Optimization block, so that all settings from the original model are carried over to the optimized one. After entering the material property values from Tables 1, 2 and 3, the software reports the model mass as 148 g. The MultiZone tool was used to generate the model mesh, producing a structured, predominantly hexahedral mesh. The fixed-support region was defined on the elements around the screw holes where the arm connects to the frame. The location of the applied force was defined on the flange, on the elements around the motor-mount screws, on one side only.

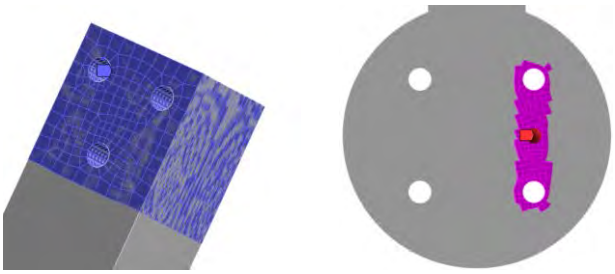


Figure 2. Fixed support – left, force location – right

The static structural analysis yields a maximum deformation of $\delta = 1.403$ mm, which is about 10% lower than the value obtained analytically. This is expected, since the analytical model is a simplified cantilever beam that deviates from the full 3D finite element model — it does not account for the Poisson's effect, the local stiffness of the cross-section, the stress distribution across the cross-section, etc. The equivalent (von Mises) stress amounts to $\sigma_{vM} = 8.466$ MPa. This is a concentrated stress present in a small region around the arm's fixed support.

When selecting the region for topology optimization, the areas where boundary conditions and loads are applied are excluded. The optimization method used here is Level Set, since the Mixable Density method does not accept the deflection criterion as a valid constraint, while the Density Based method is deprecated. This applies to the software version used in this work, Ansys Workbench 25, and does not imply that the other two methods cannot be verified in other versions. For the Level Set type, the C-MFD algorithm was selected — an enhanced version of the classical MFD algorithm, offering significantly faster solution convergence and more efficient mesh handling during the optimization process. The solver was set to run at most 500 iterations, with a convergence accuracy of 0.1%. An objective was defined for the optimized model to retain 65-70% of the original mass while maintaining

the minimum allowable stiffness, as well as a maximum deformation of 1.59 mm, preserving a safety factor of 1.1.

The geometry obtained directly from topology optimization is rarely suitable for manufacturing by conventional processes, which is why it is nowadays most commonly associated with additive manufacturing. For this reason, additional manufacturing constraints are introduced into the optimization process itself: a minimum (and maximum) member thickness, and a maximum overhang-angle constraint characteristic of FDM technology. A minimum member size of 4 mm was imposed for the optimized model, in order to prevent the formation of structurally meaningless, thin branches (or isolated, unconnected elements) that would be neither printable nor load-bearing. The default value of 45 degrees was used for the overhang angle, in the direction in which the model is printing (Z-axis). This setting ensures that the resulting shape can be printed without requiring excessive support structures.

stress state. To simplify the model for printing, the edges of the I-profile were straightened, and small mesh fragments were removed from the motor mount. The segment removed between the flange and the central part was restored, so that this refined model retains a longitudinal axis of symmetry. After this modification, the part proceeds to the final analyses.

For the refined model, the force was assigned to the elements around the joint with the motor mount. The mass of this new model version is 106 g, about 3% higher than that of the optimized model. The maximum deflection is 1.58 mm, while the maximum von Mises stress is 9.225 MPa, occurring concentrated in the regions around the fixed support.

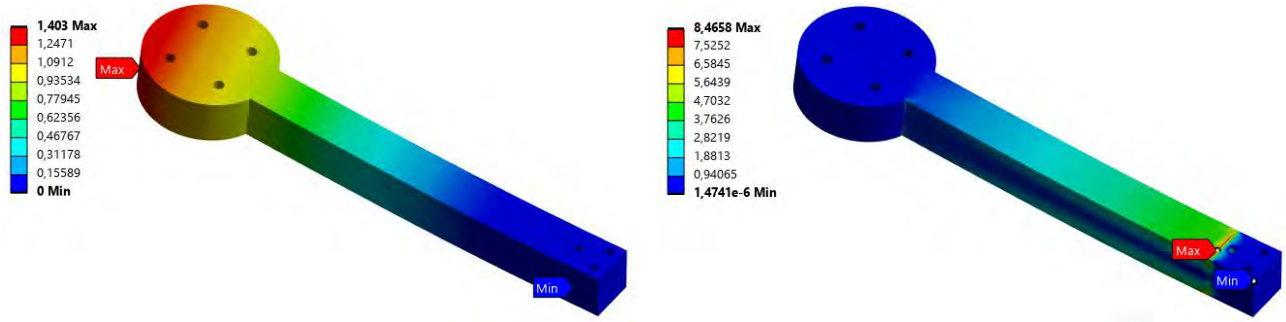


Figure 3. Total deformation [mm] – left, von Mises stress [MPa] on the original model – right

Figure 4. Optimized model – left, refined model – right

Following the optimization, material was removed from the flange, while enough material was left around the screw holes to allow the nuts to be fully threaded on. On the right-hand side of the arm (viewed from above), at the transition from the flange to the central part of the arm, material was removed, since Figure 3 shows that the right side of the arm experiences a lower load, due to the force being positioned on the left half of the motor mount. The model's mass is 103 g, which is 30% less than the original part; the maximum deformation is 1.53 mm, while the maximum normal stress is 5.284 MPa. In its central part, the optimized model has a non-symmetric I-profile cross-section, a shape well suited to resisting the combined

3. CONCLUSION

The optimized model successfully achieves what was intended: excess material has been removed, mass has been significantly reduced, and less time is required for printing. The model satisfies the maximum-deflection and equivalent-stress criteria. Following the optimization, the model's mass was reduced from 148 g to 106 g (~28%), while the maximum deflection (1.58 mm) and equivalent stress (9.225 MPa) both remain within the imposed limits, with the deflection safety factor retained at approximately 1.1. It has been shown that the Level-Set method can effectively optimize the model for the loading case considered here. Once the optimization is complete, the

model remains subject to free-form modeling by the designer, who prepares it for printing; this step requires modifying the model in line with the optimization recommendations from the software and with the principles of 3D printing.

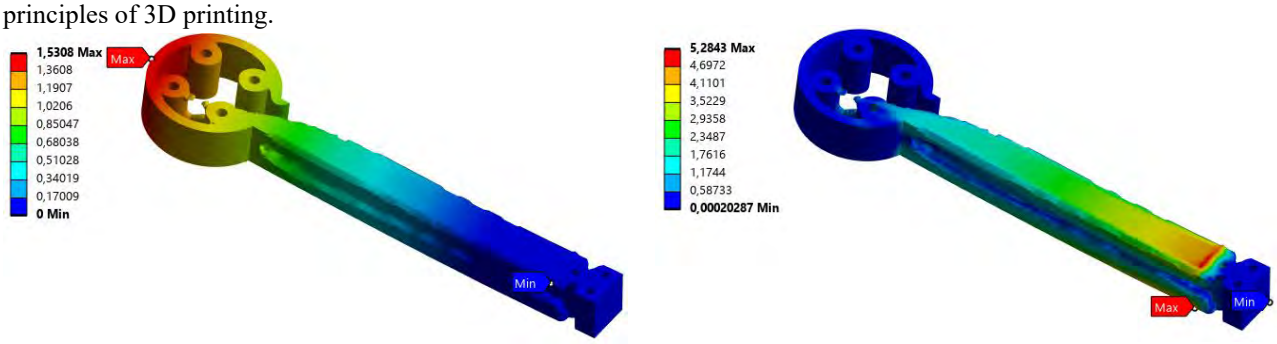


Figure 5. Total deformation [mm] – left, von Mises stress [MPa] on the optimized model – right

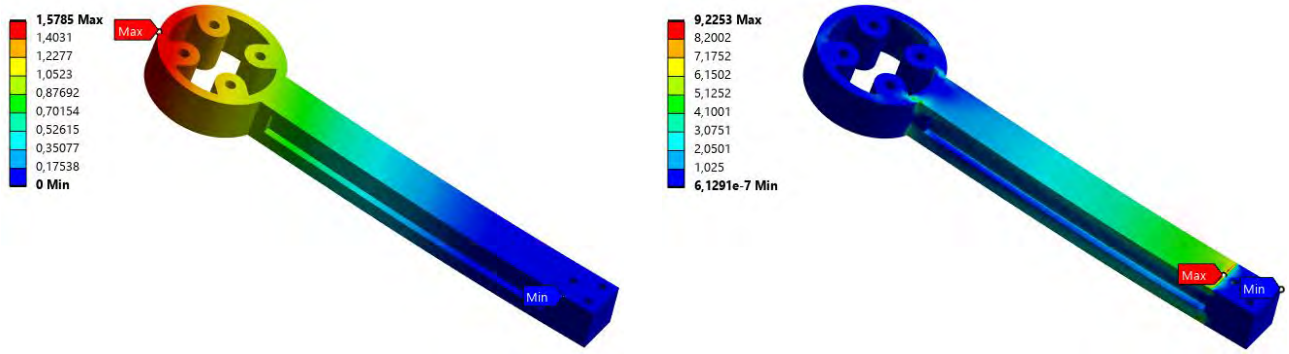


Figure 6. Total deformation [mm] – left, von Mises stress [MPa] on the refined (final) model – right

This is a drawback of this method, as it requires rework that takes up the designer's time, particularly for more complex geometries. For structural optimization purposes, it would also be worthwhile to try other approaches such as Lattice, Shape, and Topography optimization; such an approach would exceed the scope of this paper. As a continuation of this project, the printed, optimized arm needs to be fitted with the propulsion system, and its maximum deformation measured, in order to compare it with the numerical model. It is possible to consider variations in the way the arm is connected to the drone structure, so as to require fewer machine parts. It is also expected that the arm design will change as the project develops, should an alternative location be found for the ESC controllers, which currently occupy space along the drone arms — not an ideal solution.

References

[1] – D. Kotarski, P. Piljek, M. Pranić et al.: A Modular Multirotor Unmanned Aerial Vehicle Design Approach for Development of an Engineering Education Platform, Sensors 2021.

[2] – Bendsøe, M.P., Kikuchi, N. (1988). *Generating optimal topologies in structural design using a homogenization method.* Computer Methods in Applied Mechanics and Engineering, 71(2), 197–224. DOI: 10.1016/0045-7825(88)90086-2

[3] – Bendsøe, M.P., Sigmund, O. (2003). Topology Optimization: Theory, Methods, and Applications. Springer. DOI: 10.1007/978-3-662-05086-6

[4] – Sigmund, O. (2001). A 99 line topology optimization code written in Matlab. Structural and Multidisciplinary Optimization, 21, 120–127.

[5] – Lei, M., Wang, Y., et al - Numerical Homogenization Calculation of Effective Stiffness of Fused Deposition Modeling Printing Carbon Fiber Reinforced Polylactic Acid Composites, 3D PRINTING AND ADDITIVE MANUFACTURING, Volume 11, Number 5, 2024.

[6] – Wang, M.Y., Wang, X., Guo, D. (2003). A level set method for structural topology optimization. Computer Methods in Applied Mechanics and Engineering, 192(1–2), 227–246.

[7] – Zillober, C. (2002). SCPIP – an efficient software tool for the solution of structural optimization problems. Structural and Multidisciplinary Optimization. DOI: 10.1007/s00158-002-0248-5

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PROPELLER-MOTOR MATCHING FOR MINI-UAV: THRUST AND EFFICIENCY ESTIMATION BY USING THE LY-MICRO-MAX TEST STAND

MIODRAG MILENKOVIĆ-BABIĆ
Military Technical Institute, Belgrade, miodragmbm@yahoo.co.uk

BRANISLAV OSTOJIĆ
Military Technical Institute, Belgrade, branislavostojic1994@gmail.com

ALJOŠA LJUBOJA
Military Technical Institute, Belgrade, aljosabiljana@gmail.com

MARIO SILVAGNI
Department of Mechanical and Aerospace Engineering – Mechatronics Laboratory, Politecnico di Torino, Torino
mario.silvagni@polito.it

Abstract: *Unmanned Aerial Vehicle design and implementation are evolving rapidly to meet the stringent demands of civil and commercial applications. In particular, Miniature Unmanned Aerial Vehicles with a maximum take-off mass below 15 kg have seen widespread use in various sectors. Continuous technological improvements require the development of new propulsion systems in order to satisfy these new challenges. In this paper, the AXI 4130/20 Gold Line V3 brushless direct current motor, equipped with Aeronaut CAM Folding propellers, will be evaluated using a commercial-off-the-shelf test bench: the LY-10KGF Micro-Max Test Stand. The identified propeller-motor characteristics could be used as a guide for selecting the optimal propeller for small Unmanned Aerial Systems.*

Keywords: *UAV, propeller, BLDC motor, testing.*

1. INTRODUCTION

Small Unmanned Aerial Vehicles (UAVs) were initially primarily used for recreational purposes by aeromodelling enthusiasts. They subsequently found applications in the military sector, and the 21st century [1-8] has seen an expansion in their use for both military and civilian surveillance and reconnaissance missions. Today, multicopters dominate the hobbyist market due to their ease of use and Vertical Take-Off and Landing (VTOL) capabilities, which totally eliminate the need for a runway. Most of these small UAVs are powered by brushless direct current (BLDC) motors and batteries that supply energy to the UAVs. Furthermore, they can carry various types of sensors, including Electro-Optical/Infrared (EO/IR) sensors, rangefinders, weather and pollution sensors, and other specialized equipment.

For conceptual UAV design, there are numerous software tools and textbooks [9-12] available for preliminary UAV sizing and performance estimation. The accuracy of these methods is in direct correlation with the motor-engine performance data, which are often difficult to obtain or estimate accurately. In this paper, the AXI 4130/20 GOLD LINE V3 BLDC motor is evaluated alongside its propeller to estimate static propeller performance. The motor characteristics were previously validated and documented in [13]. According to the manufacturer data [14], this motor is designed for large sailplanes weighing up to 7 kg and sports aerobatic models up to 5 kg. The recommended propellers for various UAV configurations are 15×10, 15×12, 18×10, 18×8, 18.5×12, and 16×10 inches. The specifications of the AXI 4130/20 BLDC motor used in this study are summarized in Table 1.

Table 1. Technical specifications of the AXI 4130/20 GOLD LINE V3 BLDC motor

Number of cells	6-8s Li-Poly
RPM/V	305
Maximal efficiency	90%
Maximal efficiency current	18-40 A (>84%)
No load current	1.1 A
Current capacity	56 A/60 s
Internal resistance	43 mΩ
Dimension (diameter x length)	49.8 x 64.4 mm
Shaft diameter	6 mm
Weight	0.410 kg
Maximal Power	1650 W

Corresponding author: M. Milenković-Babić;
E-mail: miodragmbm@yahoo.co.uk
DOI: 10.5937/OTEH26014M

As shown in Table 1, the maximal motor power is 1650 W, which is sufficient to power small UAVs of even greater weight (up to 15 kg). However, for such higher weights, the take-off is usually performed using an external launch mechanism. The most commonly used launchers are pneumatic and mechanical systems [15, 16], favored for their reliability and compact design. To evaluate UAV performance, propeller data must be analyzed. During the take-off phase, the sufficiency of the control authority must be considered [13, 17, 18]. In order to select an adequate propeller, or to narrow down the number of potential candidates, the data from [13, 19, 20] should be used as a guide.

In this paper, the AXI 4130/20 Gold Line V3 BLDC motor, equipped with Aeronaut CAM folding propellers, is evaluated under static conditions using a commercial-off-the-shelf test bench: the LY-10KGF Micro-Max Test Stand. The results presented in this paper, together with the motor characteristics identified in previous work [13], serve as a valuable basis for optimizing the propeller–BLDC motor combination across various UAV configurations, and are especially useful for multicopter design. By analyzing measured performance parameters specifically: torque, power, and thrust, this dataset enables the selection of the optimal propeller for a defined UAV mission.

2. EXPERIMENTAL TESTING OF MOTOR-PROPELLER PERFORMANCE

In previous work, the AXI BLDC motor characteristics were defined at various supply voltages [13]; therefore, that procedure is not repeated here in detail. Typically, the battery supplies energy to the electronic speed controller (ESC), which converts direct current (DC) into a three-phase trapezoidal supply that drives the BLDC motor. The characteristics of the battery and ESC can be assumed constant in the presented experiment, as they primarily depend on the manufacturing and production processes. Based on this assumption, and the fact that all propellers are tested under identical operating conditions, any differences in the measured results are directly correlated with the specific propeller types.

In order to identify the characteristics of the different propellers, testing must be performed at defined supply voltages. In the presented analysis, an 8S battery configuration with a nominal operating voltage of 29.6 V was used. The characteristics of the AXI 4130/20 Gold Line V3 BLDC motor, based on the data from [14], are marginally different from the AXI 4130/20 Gold Line BLDC motor identified in previous work [13]. Therefore, the identified characteristics of the AXI 4130/20 Gold Line BLDC motor shown in Figure 1 will be used, with a sufficiently small error that will not affect the final results.

Propeller performance was tested by using the LY-10KGF Micro-Max Test Stand [21], shown in Figure 2. The system is designed for static thrust and torque measurements of small propulsion units. It enables the simultaneous measurement of thrust force, torque, rotational speed (RPM), current, and voltage. These measurements allow for the accurate determination of power and efficiency.

The calibration of the test bench was performed before the experimental measurements according to the manufacturer's recommendations in order to zero the thrust and torque sensors. The tests were conducted under laboratory-controlled conditions with a constant ambient temperature of 20°C. This constant temperature ensures a consistent air density throughout all experimental measurements.

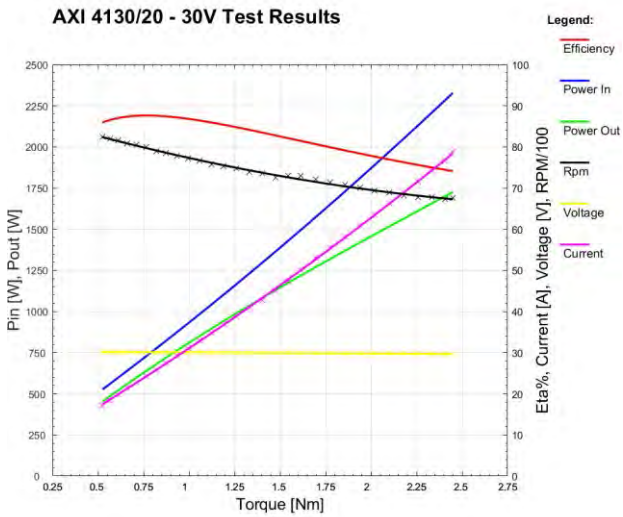


Figure 1. AXI 4130/20 test results at supply voltages of 30 V: Experimental measurements (markers) and identified (solid lines)

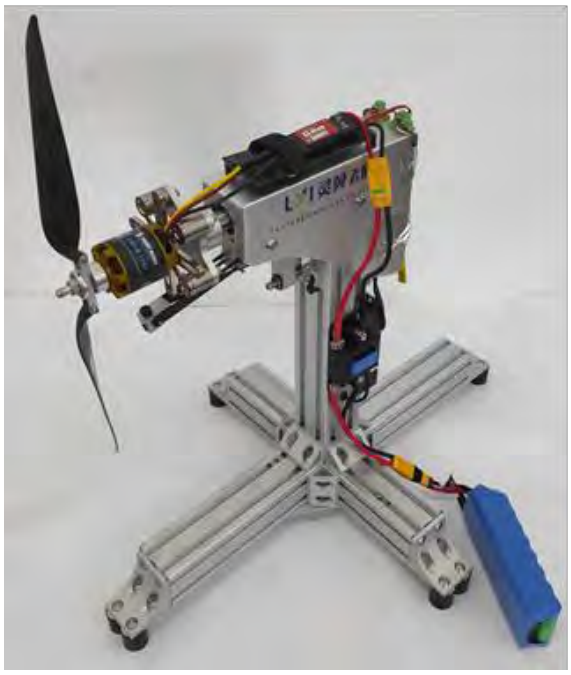


Figure 2. LY-10KGF Micro-Max Test Stand

The experimentally obtained results for motor–propeller torque, thrust, input and output power, and motor efficiency for five Aeronaut CAM Folding propellers (16×10, 16×13, 17×11, 17×13, and 18×11) at 100%, 90%, 80%, 70%, 60%, and 50% throttle settings are presented in Table 2. These measurements were conducted using a

propeller spinner with a diameter of 52 mm.

Table 2. Measured AXI 4130/20 GOLD LINE V3 BLDC motor-propeller characteristics

Propeller: 18×11 in						
Throttle [%]	**M [Nm]**	**F [N]**	**RPM**	**P_{in} [W]**	**P_{out} [W]**	**η_{motor} [%]**
100	1.72	55.31	6645.72	1429.51	1198.41	0.84
90	1.68	54.68	6588.19	1380.35	1160.84	0.84
80	1.30	35.16	6008.36	1018.76	820.00	0.80
70	0.86	28.38	4833.80	525.91	436.70	0.83
60	0.63	20.83	4125.71	333.27	271.75	0.82
50	0.32	11.59	3147.59	139.51	104.35	0.75
Propeller: 17×13 in						
Throttle [%]	**M [Nm]**	**F [N]**	**RPM**	**P_{in} [W]**	**P_{out} [W]**	**η_{motor} [%]**
100	1.89	43.72	5981.44	1532.10	1186.29	0.77
90	1.85	44.14	6265.80	1486.76	1212.56	0.82
80	1.50	35.72	5699.28	1089.52	897.94	0.82
70	1.28	30.08	5250.82	850.27	701.21	0.82
60	0.97	23.61	4629.02	578.32	471.42	0.82
50	0.69	17.16	3992.73	357.86	287.70	0.80
Propeller: 17×11 in						
Throttle [%]	**M [Nm]**	**F [N]**	**RPM**	**P_{in} [W]**	**P_{out} [W]**	**η_{motor} [%]**
100	1.89	51.12	6115.75	1550.38	1210.94	0.78
90	1.84	50.21	6408.50	1504.64	1235.18	0.82
80	1.47	41.57	5822.24	1091.48	898.35	0.82
70	1.23	35.18	5362.55	837.52	691.64	0.83
60	0.93	26.76	4716.51	564.15	461.73	0.82
50	0.66	19.09	4081.48	356.75	280.92	0.79
Propeller: 16×13 in						
Throttle [%]	**M [Nm]**	**F [N]**	**RPM**	**P_{in} [W]**	**P_{out} [W]**	**η_{motor} [%]**
100	1.88	40.43	6143.73	1555.55	1210.58	0.78
90	1.83	39.67	6426.01	1510.16	1234.11	0.82
80	1.49	31.93	5868.04	1114.88	914.94	0.82
70	1.22	27.55	5380.92	835.81	688.54	0.82
60	0.93	21.78	4754.31	563.89	460.61	0.82
50	0.67	15.79	4046.01	353.19	282.12	0.80
Propeller: 16×10 in						
Throttle [%]	**M [Nm]**	**F [N]**	**RPM**	**P_{in} [W]**	**P_{out} [W]**	**η_{motor} [%]**
100	1.53	41.71	6618.74	1308.31	1061.39	0.81
90	1.49	40.75	6939.03	1269.55	1082.93	0.85
80	1.11	33.60	6322.33	861.38	736.79	0.86
70	0.91	28.34	5779.34	645.28	549.09	0.85
60	0.68	21.38	5050.63	431.12	359.69	0.83
50	0.48	15.51	4297.77	271.40	216.57	0.80

It must be noted that propeller efficiency could not be defined, as static performance was evaluated, and under such conditions, propeller efficiency is mathematically undefined (since output power, as the product of thrust and airspeed, is zero during static tests).

3. DISCUSSION OF THE RESULTS

First, the characteristics of the AXI 4130/20 BLDC motor were identified to provide a solid basis for an accurate estimation of motor-propeller performance.

From an initial analysis of the data in Table 2, it is evident that propellers with a 13-inch pitch (17×13 and 16×13) are less efficient than propellers with a smaller pitch at maximum throttle settings. Based on the identified data, it can be concluded that the optimal throttle for a UAV using a runway for take-off is not necessarily the maximum throttle, according to the recommendation in [18]. The AXI 4130/20 BLDC motor equipped with the 16x10 in propeller exhibits the highest efficiency within the throttle range of 60% to 90%. This high efficiency likely explains why this specific propeller is suggested for sailplane UAVs weighing approximately 7 kg.

The data from Table 2 indicate that the 18×11 in propeller provides the highest static thrust, which enables the shortest take-off distance from a runway. Additionally, it features the best thrust-to-torque ratio, making it an ideal choice for the take-off phase, especially for unconventional UAV designs [8]. Additionally, the identified static thrust and the remaining motor–propeller data can be used to optimize the hover capability of UAVs designs.

3. CONCLUSION

The conducted experimental and numerical investigations including motor parameter identification from previous work, as well as static thrust and torque measurements have provided comprehensive data on motor–propeller performance across various configurations.

The results obtained from the motor–propeller characterization (Table 2) clearly indicate that selecting the optimal propeller for a UAV is a complicated task. It must account for maximum thrust, the torque-to-thrust ratio, propeller rotational speed, and the UAV's maneuvering capabilities.

Based on the results presented in Table 2, the 18×11 in propeller demonstrated the highest static thrust force and the most favorable thrust-to-torque ratio. These data indicate that it represents an optimal choice for UAVs that require special emphasis on the take-off flight phase. Furthermore, the results indicate that the AXI 4130/20 V3 BLDC motor, when equipped with the 16×10 in propeller, achieves the highest efficiency. This explains why this specific propeller is recommended as the optimal choice by the motor manufacturer.

Acknowledgments

This research was funded by the Military Technical Institute (VTI) Belgrade and partly by the Ministry of Science, Technological Development and Innovation of Republic of Serbia (Project number 451–03–34/2026–03/200325).

References

[1] ANTONIĆ, V., TRIFKOVIĆ, M., MOLOVIĆ, V., MILENKOVIĆ-BABIĆ, M.: Small Fixed-Wing UAV Precision Aerial Drop Capability Development, Scientific Technical Review, 74(2) (2024) 65-70.

[2] MARKET RESEARCH FUTURE: Global Unmanned Aerial Vehicles Market Overview, Available online: https://www.marketresearchfuture.com/reports/unmanned-aerial-vehicle-uav-market-806, (2024).

[3] JANES: All the World's Aircraft: Unmanned, 2024-2025 Yearbook, Janes Information Services, (2024).

[4] MILENKOVIĆ-BABIĆ, M., DOVATOV, B., OSTOJIĆ, B., ANTONIĆ, V., TRIFKOVIĆ, M., IVKOVIĆ, D.: Tactical UAV flight performance estimation and validation, Defence Science Journal, 75(1) (2025) 19-26.

[5] SAEED, A.S., YOUNES, A.B., CAI, C., CAI, G.: A survey of hybrid Unmanned Aerial Vehicles, Progress in Aerospace Sciences, 98 (2018) 91-105.

[6] MARIN, M., MIROSAVLJEVIĆ, P.: Analysis of the performance and kinematics of the movement of UAV, FME Transactions, 51(4) (2023) 627-636.

[7] MILENKOVIĆ-BABIĆ, M., IVKOVIĆ, D., OSTOJIĆ, B., et al.: Flying Wing Conceptual Design and Flight Testing, FME Transactions, 52(4) (2024).

[8] MILENKOVIĆ-BABIĆ, M., OSTOJIĆ, B., DOVATOV, B., et al.: Control derivative estimation of X-tail control surface design, Aerospace Science and Technology, 163 (2025), DOI: 10.1016/j.ast.2025.110214.

[9] GUDMUNDSSON, S., General aviation aircraft design: Applied methods and procedures. Elsevier Inc, 2014.

[10] NAPOLITANO, M., Aircraft Dynamics: From Modeling to Simulation, John Wiley and Sons, 2011.

[11] RAYMER, D., Aircraft Design: A conceptual Approach, Fifth edition, AIAA education series, 2012.

[12] https://openvsp.org/.

[13] MILENKOVIĆ-BABIĆ, M., OSTOJIĆ, B., RISTIĆ, N., DOVATOV, B., et al.: Compensation of Motor Torque Asymmetry during the Take-off Phase of an X-tail unmanned aerial vehicle, FME Transactions,

54(2) (2026).

[14] www.modelmotors.cz/product/detail/276/.

[15] BEK-UZAROV, M., NENADIĆ, I., RAŠIĆ, S.: Design and verification of a pneumatic launch system for a single-use UAV with a warhead, 11th International Scientific Conference on Defensive Technologies – OTEH 2024, (2024) 60-64.

[16] NOVAKOVIĆ, Z., MEDAR, N.: UAV launch systems, Armament Data – Fact Sheets, Military Technical Institute, Belgrade, 38(156) (2015). (In Serbian).

[17] STOJAKOVIĆ, P., RAŠUO, B.: Single propeller airplane minimal flight speed based upon the lateral maneuver condition, Aerospace Science and Technology, 49 (2016) 239-249.

[18] STOJAKOVIĆ, P., VELIMIROVIĆ, K., RAŠUO, B.: Power optimization of a single propeller airplane take-off run on the basis of lateral maneuver limitations, Aerospace Science and Technology, 72 (2018).

[19] APC propeller performance data. Available online: https://www.apcprop.com/technical-information/performance-data.

[20] UIUC propeller data site. Available online: https://m-selig.ae.illinois.edu/props/propDB.html.

[21] https://www.wingflyingtech.com/data/upload/20210604/60b9ed2bd9a3f.pdf.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ESSENTIAL TECHNICAL KNOWLEDGE AND SKILLS FOR FUTURE MILITARY AND CIVIL PILOTS IN THE REPUBLIC OF SERBIA

SLAVIŠA VLAČIĆ
Air Academy, Belgrade vlacic.s@vakademija.edu.rs

ALEKSANDAR KNEŽEVIĆ
Military Academy, University of Defense, Belgrade aleksandar.knezevic@va.mod.gov.rs

NEBOJŠA RADOJKOVIĆ
204. Air Force Brigade, Serbian Armed Forces, Belgrade bojsha79@hotmail.com

Abstract: *This paper describes and presents the essential technical knowledge and skills for future military and civilian pilots. The observations presented on this topic are based on current legal regulations, experience, training standards, as well as the assumed direction of aviation technology development and the approach to its operation. Special attention is dedicated to the similarities and differences in the types and levels of technical knowledge required for future military and civil pilots, and their potential harmonization, keeping in mind the modernization processes of both military and civil aviation with advanced aircraft technology in the Republic of Serbia.*

Keywords: *technical knowledge, skills, pilot, modernization process.*

1. INTRODUCTION

In the Republic of Serbia, aviation can be categorized into military and civil sectors, encompassing both military and civilian pilots. The civil aviation sector is further divided into commercial and non-commercial segments. A notable characteristic of aviation in Serbia today is the ongoing programs for extensive procurement and modernization of both military and civilian aircraft.

These programs are changing the requirements placed before the entire aviation personnel, especially pilots, both military and civil, who are required to master and operate these assets. In the initial phase of receiving aviation hardware, the focus is primarily on pilots who are completing conversion training and adapting to the operation of new aircraft. However, to complete conversion training high-quality, efficiently, and safely-especially when dealing with a significant technological leap-a high level of previously acquired technical knowledge and skills is required, gained through formal education, flight practice, or complex training programs. These requirements can be defined uniformly for both military and civil pilots in the Republic of Serbia, although the path to acquiring these knowledges and skills differs both chronologically and methodologically. Furthermore, the standards and regulations governing the attainment of the required level of core technical knowledge and skills for military and civil pilots in the Republic of Serbia also differ. This paper details specific aspects of the process of acquiring technical knowledge and skills for military and civil pilots, as well as the changes and needs expected in the upcoming period. It is precisely these changes and needs that will define the core technical knowledge and skills of future military and civil pilots in the Republic of Serbia. Predictions regarding specific knowledge and skills are also highlighted in this paper.

2. CATEGORIES OF MILITARY AND CIVIL PILOTS

When viewed as professional categories, military and civil pilots differ in numerous ways, although both are engaged in aircraft operation and air navigation. The philosophy and execution of their tasks usually differ significantly. The extent of these differences depends on the aviation subcategory. Simplified, military aircraft in the Republic of Serbia are divided into combat and non-combat platforms, as well as fixed-wing aircraft and helicopters. Generally, we can speak of fighters, fighter-bombers, transport aircraft, attack helicopters, multirole helicopters, and transport helicopters.

If we consider civil aviation in the Republic of Serbia, attention is focused on the national carrier, Air Serbia, along with several smaller companies and training centers engaged in commercial operations, aerial work (general aviation), and advanced pilot training. A significant aviation organization is also the Helicopter Unit of the Ministry of Internal Affairs of the Republic of Serbia (equipped with multirole and transport helicopters). For these exact reasons, we have categorized pilots by sector (*civilian*) rather than by primary purpose (*commercial*), because despite the designation of the license type (*Commercial Pilot License*), not all pilots or aviation organizations in the Republic of Serbia are engaged in commercial air transport.

From this, a wide diversity of aircraft and their

Corresponding author: S. Vlačić;
E-mail: vlacic.s@vakademija.edu.rs
DOI: 10.5937/OTEH26015V

operational roles can be observed, implying different types of training and knowledge required for the pilots operating these platforms.

3. MODERNIZATION PROGRAMMES IN MILITARY AND CIVIL AVIATION IN THE REPUBLIC OF SERBIA

To define the core technical knowledge and skills of future military and civil pilots, it is necessary to identify the modernization processes in military and civil aviation over the past decade, as this reveals the technological shift from which the need for new knowledge and skills arises. The Serbian Air Force (SAF) practically initiated its modernization programs in 2012 with the introduction of the Lasta basic trainer aircraft, featuring a digital glass cockpit. The following year, in 2013, the national carrier Air Serbia commenced operations, opting from the start for European passenger aircraft manufactured by Airbus rather than American aircraft from Boeing. Following modernization trends in Serbia, the procurement package also initiated the modernization of the helicopter fleet of the Ministry of Internal Affairs (MoI) Helicopter Unit, which in the first phase received H145 helicopters identical to those of the SAF, followed by other specialized aircraft. An overview of aircraft procurements and modernizations for the SAF by year, role, and type is given in Table 1.

Table 1. Procurements and modernizations for the SAF

Airplane/ Helicopter	Role	Acquisition Year	Replace for:
Lasta	Basic Training	2012	Utva-75
MiG-29 SM	Multirole	2021	MiG-29 (izd.9-12)
C-295	Transport	2024	An-26
Dassault Rafale F4	Multirole	2028	MiG-29
Mi-17	Transport	2016	Mi-8
H145M	Multirole	2019	Sa-341/342
Mi-35M	Combat	2019	Sa-341/342 (Gama)
Mi-35P (Ex-Cypriot)	Combat	2023	Sa-341/342 (Gama)

An overview of aircraft procurements for Air Serbia by year and type is given in Table 2.

Table 2. Aircraft procurements for Air Serbia

Airplane	Operator	Acq.Year	Replace for:
Airbus A319	AirSerbia	2013	Boeing 737-300
Airbus A320	AirSerbia	2014	Boeing 737-300
Airbus A330-200	AirSerbia	2016	DC-10
ATR 72-600	AirSerbia	2023	ATR 72-500
Embraer E-195	AirSerbia	2024	-

An overview of helicopter procurements for the MoI Helicopter Unit by year and type is given in Table 3.

Table 3. Helicopter procurements for the MoI

Helicopter	Operator	Acq. Year	Replace for:
H145M	MoI	2019	Sa-341/342, Bell 206
H215/225 Super Puma	MoI	2022	Bell 212, S-76
Ka-32A11BC	MoI	2022	-

As another major operator, the ATO Air Academy in Vršac acquired a completely new fleet of Cessna 172S aircraft equipped with Garmin G1000 glass cockpit avionics in 2012.

What can be observed from the previously presented tables are the following:

- Virtually the entire civil and military fleet state-operated within the Republic of Serbia has been replaced with new or modernized aircraft, regardless of the specific user.
- All new aircraft feature digital glass cockpits, while modernized platforms and older types have been upgraded to semi-digital cockpits.
- The vast majority of new fixed-wing aircraft and helicopters were manufactured in Western countries.
- There is a clear strategic tendency toward relying on European manufacturers (Airbus and Dassault).
- Modernization of older Russian/Soviet-manufactured aircraft included Western/ICAO-standard avionics systems for navigation and air traffic control, including GPS/GNSS, civil ATC-compliant VHF radios, civilian-compatible transponders, and VOR/ILS capability.
- Newer Russian types, such as the Kamov Ka-32A11BC, were specifically designed to meet Western airworthiness standards. In 2009, it officially received its EASA Type Certificate. Considering these facts, technical knowledge and skills must be adapted more fully to this operating environment. Depending on the user and operator, education and training systems should be adjusted in line with directives issued by the regulatory bodies of the manufacturing countries. In this case, the relevant manufacturers are European, and their operations are largely governed by EASA (European Union Aviation Safety Agency) rules and regulations. Failure to recognize and apply these directives increases training costs, extends training duration, and may compromise flight safety.

4. THE WAY OF ACQUIRING PILOTS' KNOWLEDGE AND SKILLS

In the Republic of Serbia, there are generally three main educational and training streams through which individuals enter professional military and civil flight

duties as pilots.

The first and most complex pathway is the *Military Aviation* study program at the Military Academy, University of Defence in Belgrade [1]. This program is distinctive because it aligns with Bologna Process standards while also meeting Air Force operational requirements. It educates Military Academy cadets—undergraduate students in military terminology—for roles as fixed-wing or helicopter pilots, as well as mechanical engineers. The undergraduate Military Aviation program lasts four academic years and carries 240 ECTS credits. Upon completion, cadets earn a Bachelor of Science degree in Mechanical Engineering and receive a Diploma Supplement certifying their competencies as professional Air Force pilots.

The curriculum is divided into two main parts. The first phase, lasting six semesters, is conducted at the Military Academy headquarters in Belgrade. The primary goal of this phase is theoretical preparation for upcoming flight training in training squadrons. The first group of courses consists of theoretical, methodological, and general scientific subjects, while the second group comprises professional and engineering courses. Study of the second group occurs during the 3rd academic year (5th and 6th semesters) including subjects such as General Navigation, Radio Navigation, Communications, Principles of Flight (Aerodynamics/Flight Mechanics), Aircraft General Knowledge (Aircraft Construction), Meteorology, Powerplants (Aircraft Engines), Instrumentation (Flight Instruments and Systems), Mass and Balance, Human Performance and Limitations, Operational Procedures, and Air Law. A portion of these subjects is tailored specifically for future helicopter pilots. Due to international standardization, most listed courses are compatible in scope and content with similar courses taught at aviation universities or ATOs worldwide. Military-specific subjects are not evaluated in this paper. Flight training is executed in training squadrons during the final year of study.

The second stream, which has emerged over the last five years, is offered through the Aviation Academy (AA) in Belgrade and its ATO (Approved Training Organization) in Vršac. In close cooperation with the Academy of Applied Studies Polytechnic, AA educates pilots through the *Air Transport* study program over three academic years (180 ECTS). The main goal is to educate professional transport engineers with the knowledge, competencies, and practical skills required to perform engineering tasks in air transport, as well as to acquire the theoretical qualifications necessary for obtaining an Airline Transport Pilot License (ATPL) with an Instrument Rating (IR), in accordance with EASA Flight Crew Licensing regulations (Part-FCL) [2]. The program prepares professional engineers for professional pilot licenses, including the frozen Airline Transport Pilot Licence – Aeroplane (ATPL(A)). The curriculum, comprising compulsory and elective courses, aligns with requirements prescribed by EASA Part-FCL regulations. It includes 13 core theoretical subjects: Aircraft General Knowledge, Instrumentation, Performance, Mass and Balance, Flight Planning and Monitoring, General Navigation, Radio Navigation, Communications, Principles of Flight, Meteorology, Human Performance and Limitations, Operational Procedures, and Air Law. According to EASA requirements, the theoretical ground school scope comprises a minimum of 750 hours. Flight training is conducted at the ATO, commencing in the 4th semester.

The third stream for obtaining licenses and becoming a professional pilot in the Republic of Serbia is through independent ATO facilities approved for CPL/ATPL training. This process is fully grounded in EASA regulations and Part-FCL rules. Practically, the theoretical syllabus and flight training are identical to stream two, but without the academic degree components. This pathway can be shorter by at least a year, which is why many aspiring pilots opt for this route.

Although an informal industry view once favored requiring all commercial pilots to hold at least a Bachelor's degree, this requirement was never incorporated into aviation regulations. It remains mandatory only for military pilots and Ministry of Interior pilots, for whom a university degree is an employment prerequisite. Commercial airlines and flight training centers may employ pilots and instructors without a Bachelor's degree, although hiring practices differ between the United States and Europe. Major U.S. airlines often prefer candidates with a college degree during competitive recruitment periods, while European airlines place greater emphasis on demonstrated competencies and ATPL performance. A mandatory global degree requirement was never adopted, partly because of persistent pilot shortages. Nevertheless, when candidates have comparable flight competencies, a formal university degree can still provide a competitive advantage.

After obtaining their licenses, candidates from streams two and three are subject to a multi-stage selection process before entering an airline or operator, designed to test core competencies (such as those defined in ICAO and EASA frameworks). This process usually consists of:

- **Screening & Technical Assessment:** Evaluating theoretical knowledge and mental math capabilities.
- **Simulator Assessment:** Usually conducted in a Flight Synthetic Training Device (FSTD) / Flight Simulator. It serves as an operator's tool to evaluate manual hand-flying skills, instrument proficiency, raw-data navigation, and elements of Crew Resource Management (CRM).
- **HR & Competency Interview:** Assessing communication skills, adaptability, teamwork, and alignment with company culture.

Some of these core competencies are also evaluated in military pilots before initiating conversion training to a new aircraft type.

Considering the aforementioned points, all three entry streams feature a highly similar theoretical syllabus

defined by EASA regulations.[1] The detailed ATPL theoretical knowledge syllabus and its Learning Objectives (LOs) are defined in the Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Part-FCL [3]. The syllabus is subject to periodic updates. The structural framework and core syllabus currently used across Europe—often referred to as the "2020 Syllabus"—was introduced by ED Decision 2018/001/R. This major regulatory decision was enacted eight years ago.

5. TECHNICAL KNOWLEDGE AND SKILLS

When analyzing the EASA ATPL theoretical syllabus followed by most aviation training organizations in Serbia, it becomes evident that technical knowledge serves as the underlying theoretical foundation explaining why aircraft systems function and how physical laws govern aircraft behavior.

In the EASA ATPL theoretical program, this technical knowledge is systematically developed across a 13-subject syllabus, which can be categorized into distinct subject areas:

1. **Aircraft Architecture & Physics:** Provides the fundamental engineering, mechanical, and physical concepts behind how an aircraft flies and operates. It comprises Principles of Flight, Aircraft General Knowledge (Airframe and Systems, Electrics, Powerplant), and Instrumentation.
2. **Navigation, Flight Planning & Operations:** Covers the quantitative, mathematical, and spatial knowledge required to navigate the aircraft safely and efficiently. This group includes General Navigation, Radio Navigation, Flight Planning and Monitoring, Mass and Balance, and Performance.

While technical knowledge represents theoretical understanding, technical skills reflect how a pilot executes tasks in the cockpit environment. Under internationally recognized standards, a pilot's technical skill set is divided into four main functional areas, as presented in Table 4.

Table 4. Pilot's technical skill set

Areas	Skills
Aircraft Handling & Control	Manual Precision Energy & Aerodynamic Management Upset Prevention and Recovery (UPRT)
Flight Path & Automation Management	FMS/Autopilot Operation Automation Mode Awareness Graceful Degradation
Systems Management & Troubleshooting	Aircraft System Operations Abnormal & Emergency Procedures Failure Diagnosis In-Flight Mental Calculations
Navigation & Environmental Execution	Chart Reading & PBN Weather Analysis Regulatory & SOP Compliance

Additionally, military authorities typically possess specialized knowledge and skills, which they express differently. An example of how this knowledge and these skills are integrated into military flight training, and the licensing process can be seen in the case of the Czech Air Force [4].

There are some of the skills that graduates of the study specialization of military pilots have to acquire:

- Aircraft or helicopter commanders have professional knowledge at the level of the requirements of the ATPL (air transport pilot – internationally recognized standard for military and civilian pilots) professional aviation course for aircraft or helicopters and other professional knowledge necessary to obtain the appropriate military license (Aircraft or Helicopter Commander).

Graduates will utilize the theoretical and professionally oriented knowledge gained during their studies to perform the duties of an operator within a relevant military aviation system. Graduates will also:

- apply knowledge of the requirements of standards, regulations, and laws in professional activities,

- use the assigned aeronautical equipment, technology, or system to the required professional standard, including maintaining the required level of efficiency and safety,

- use aeronautical radio communication in English,

- use aeronautical equipment in the performance of assigned tasks, observing the principles of safety and efficiency of air traffic,

- evaluate information, draw conclusions, and make decisions to accomplish the assigned tasks according to the type of air force to which they are assigned,

- assess the implications of the use of weapon systems of the airborne equipment operated.

According to military aviation regulations, every student pilot must first undergo theoretical type training, the so-called material part of the training process. In this course, they become familiar with the aircraft they will be flying. Subsequently, together with the flight instructors, they undergo special purpose training (specific emergency procedures, principles of piloting techniques, important landmarks in the vicinity of the aerodrome, radiotelephone procedures, and correspondence used at the home aerodrome).

Graduates with the appropriate professional knowledge and skills at the level of the internationally recognized standard for the air transport pilot (ATPL) professional flying course will meet the requirements for the relevant military license.

Previously mentioned technical knowledge and skills can be accepted in the case of Serbian military pilots, too.

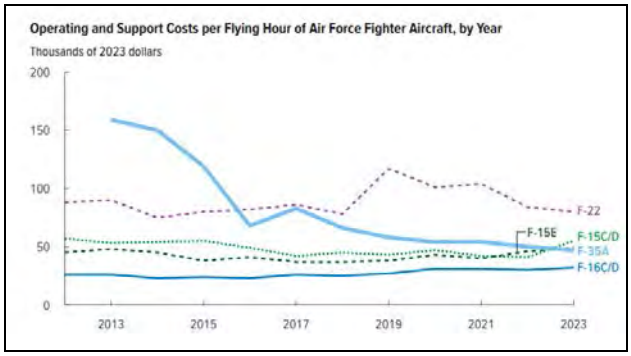
[1] The Republic of Serbia accepted EASA rules and incorporated them into its legal order since 2006 when Serbia signed the (ECAA Agreement) in Luxembourg.

But, the Serbian Air Force (SAF), however, has not yet adopted EASA standards to a sufficient extent, despite the fact that the majority of its recently acquired aircraft originate from European Union member states where EASA regulations are applied. This may create a need for additional costly conversion training and subsequent standardization, ultimately reducing personnel availability and operational readiness.

Also, when we consider the training of military pilots, it is necessary to take into account some other specificities.[2]

The specificity of military pilot training for operating modern multirole fighter aircraft is primarily reflected in the high operational and support costs (Operational & Support Costs). The cost per flight hour (CPFH) for this type of aviation can reach as much as USD 90,000. For example, during the 2023 fiscal year, the lowest cost per flight hour within the United States Air Force (USAF) was recorded for the F-16C/D, at approximately USD 30,000, while the highest was associated with the F-22, at approximately USD 90,000, as shown in Table 5. [5]

Table 5. Cost Per Flight Hour (CPFH) of Various Fighter Aircraft



When the total number of flight hours required by various military flight training programs -often reaching up to 200 flight hours - is taken into account, it becomes evident that the overall cost of training military fighter pilots is exceptionally high. Consequently, it is of paramount importance that pilot candidates for supersonic aviation possess a high level of technical knowledge and practical skills as a prerequisite for effective adaptation to the cockpit ergonomics of modern fourth and fifth-generation fighter aircraft.

Particular emphasis should be placed on mastering advanced Flight Control System/Fly-by-Wire (FCS/FBW) technologies integrated with the Hands-On Throttle and Stick (HOTAS) control concept. Training in configuring and managing the glass cockpit interface across multiple multifunction displays, depending on mission requirements, is conducted using Flight Simulation Training Devices (FSTDs) and Full Mission Simulators (FMSs). Such training significantly enhances pilots' situational awareness and decision-making capabilities in highly demanding tactical environments.

Furthermore, knowledge acquired in the field of Aviation Medicine, particularly the correct application of the Anti-G Straining Maneuver (AGSM), provides a solid foundation for developing the ability to withstand high positive vertical G-loads encountered during air combat maneuvering.

Equally important is comprehensive education in airpower doctrine, fighter tactics, and the employment of air units within modern operational environments characterized by advanced electronic warfare capabilities and the growing integration of artificial intelligence into the decision-making process. These developments represent a significant institutional challenge for military flight training organizations, requiring them to develop pilots capable of maintaining high levels of situational awareness in increasingly complex combat environments.

Counter-unmanned aircraft systems (C-UAS), including operations against remotely piloted aircraft and drones, constitute another critical area that should be incorporated into fighter pilot education and treated as a dedicated component of the training curriculum. Continuous modernization of training programs and instructional methodologies should therefore remain a permanent responsibility of organizations responsible for military flight education in order to ensure that pilots acquire the technical competencies required by the rapidly evolving nature of modern air warfare.

Although merely identifying categories of technical knowledge and skills presents little conceptual difficulty, the greater challenge lies in determining how to improve the process of acquiring those competencies in an effective manner, particularly considering the increasingly important role of technology in education and training [6].

Technological development in both military and civil aviation is progressing at a significantly faster pace than many existing training syllabi can accommodate. The intervals between revisions of EASA theoretical knowledge syllabi have proven to be excessively long. One example is the growing need for more modern approaches within subjects such as Radio Navigation, Aircraft General Knowledge, and Instrumentation, with increased emphasis on GNSS, RNP AR, GBAS, LPV, navigation integrity monitoring, ARINC 429, AFDX, More Electric Aircraft (MEA) systems, Threat and Error Management (TEM), cockpit automation, cybersecurity, and the operational use of Electronic Flight Bags (EFBs).

Rapid innovation in aviation technologies, such as advanced avionics systems, automation, and digital communication tools, necessitates updated and inventive training techniques to ensure professionals can effectively handle, maintain, and manage these systems [6].

The possible way for achieving higher level of technical knowledge and skills in the future is more comprehensive usage of immersive technologies, artificial intelligence and adaptable training systems, incorporation of game-based education and remote and cooperative instruction. A possible approach is given in [7].

[2] There is a saying „Where civilian pilot training ends, military pilot training begins."

6. CONCLUSION

This paper has identified the categories of military and civilian pilots in the Republic of Serbia, as well as the aircraft they currently operate or are expected to operate in the near future. It was observed that a significant portion of both the military and civilian aviation fleets has recently been modernized through state-funded procurement programs.

The paper has also outlined the institutions and organizations responsible for pilot education and training, with particular emphasis on the technical knowledge and practical skills that pilots are expected to acquire throughout their education and training process, or through specialized training alone in certain cases.

As an associated member of the European Union Aviation Safety Agency (EASA), the Republic of Serbia recognizes and applies the European Union's regulatory framework in the field of civil aviation. Civil aviation training organizations are therefore fully aligned with EASA standards regarding the acquisition of technical knowledge and practical competencies required of future pilots.

The Serbian Air Force (SAF), however, has not yet adopted EASA standards to a sufficient extent, despite the fact that the majority of its recently acquired aircraft originate from European Union member states where EASA regulations are applied. This may create a need for additional costly conversion training and subsequent standardization, ultimately reducing personnel availability and operational readiness.

Although the identified categories of technical knowledge and skills themselves are generally not subject to significant debate, the continuous evolution of aviation technologies will inevitably present new challenges regarding the methods by which these competencies are acquired. The intervals between revisions of the EASA theoretical knowledge syllabi have proven to be excessively long. Future research should therefore focus on identifying regulatory-compliant approaches for improving the acquisition of technical knowledge and practical skills among future pilots in the Republic of Serbian while remaining consistent with both national requirements and international aviation standards.

One of the recommended ways for achieving higher level of technical knowledge and skills in the future is more comprehensive usage of immersive technologies, artificial intelligence and adaptable training systems, incorporation of game-based education and remote and cooperative instruction.

References

[1] VLAČIĆ S., KNEŽEVIĆ A., CMILJANOVIĆ B., *Model of military pilot education and training from the aspect of new aircraft acquisition,* Vojno delo, vol. 69 (4), (2017) 36-48.

[2] EASA Part FCL

[3] Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Part-FCL

[4] KORECKI Z., HOIKA T., BAUER M., SMRŽ V., *New approach to military pilot education and training in the new security environment, Transport Problems, vol. 19 (2), (2024) 137-150.*

[5] *Availability, Use, and Operating and Support Costs of F-35 Fighter Aircraft,* CBO, USA, 2025.

[6] Dincer N., Elevating Aviation Education: *A Comprehensive Examination of Technology's Role in Modern Flight Training*, Journal of Aviation, (2023) 7(2): 317-323

[7] Mosle, B.; Schaffernak, H.; Vorraber, W.; Holy, M.; Herrele, T.; Braunstingl, R.; Koglbauer, I. Towards a *More Socially Sustainable Advanced Pilot Training by Integrating Wearable Augmented Reality Devices*. Sustainability 2022, 14, 2220.

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION III

Weapon systems and combat vehicles - WSCV

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A BALLISTIC MODULE ARCHITECTURE FOR MORTAR FIRE CONTROL SYSTEMS

MILOŠ MARKOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Weapon Systems Department, mdmarkovic@mas.bg.ac.rs

ANDRIJA OSTOJIĆ
SEDEC d.o.o., Belgrade, Serbia, andrijaostojic05@gmail.com

DEJAN JEVTIĆ
University of Belgrade, Faculty of Mechanical Engineering, Weapon Systems Department, djevtic@mas.bg.ac.rs

Abstract: *This paper presents a ballistic module architecture developed for computing firing elements in mortar systems. The ballistic module application calculates firing elements within 5 seconds, significantly reducing engagement time compared to legacy manual procedures requiring 30–60 seconds. The computational engine implements six-degree-of-freedom trajectory modeling, solving the boundary value problem through an iterative shooting method. The system architecture is organized into three integrated modules, Configuration, Ballistic Module, and Communication supporting system setup, ballistic computation with real-time meteorological data, and forward observer data exchange. Beyond real-time computation, the module enables firing table generation, supports advanced engagement modes including Multiple Rounds Simultaneous Impact, and operates autonomously in both test and combat environments. Full system configuration is completed within 40 seconds, with the architecture adaptable to multiple weapon platforms without code modification. The results confirm its effectiveness as a decision-support tool for modern mortar fire control systems, combining computational speed with the flexibility required for integration into broader command and information architectures.*

Keywords: *Ballistic module architecture, Mortar systems, Fire control systems.*

1. INTRODUCTION

In modern artillery systems, ballistic modules must balance high accuracy with rapid response times [1]. Gacek and Leśnik validated 6-DOF models experimentally [2], while NATO STANAG 4355 recommends the MPMMT as a computationally efficient alternative [3]. This model enables faster calculations through yaw of repose simplifications, making it suitable for real-time FCS applications [4]. The trade-off between model fidelity and computational speed remains central to fire control system design, as engagement windows continue to shrink in modern combat scenarios. Mortar systems, with ranges up to 10–12 km, permit flexible model selection [4,5]. Chusilp et al. demonstrated 6-DOF iterative search achieving 10 m accuracy within 1-2 minutes [5,6]. However, longer-range systems require MPMMT for rapid computation [6,7], enabling firing element determination in seconds. NATO's NABK further exemplifies standardized ballistic software components [7,8]. The ability to switch between models based on operational requirements offers significant tactical advantages, allowing mortar units to maintain precision while adapting to dynamic battlefield conditions. Modern FCS integrate ballistic computation with configuration and communication [9]. Luković et al. highlight modular, hardware-independent ballistic modules [9,10], while FENIC'S supports multiple artillery platforms [10,11]. The Artillery FCS system likewise employs MPMMT over polynomial approximations [12]. These integrated environments reduce manual data entry errors and streamline targeting, enabling forward observers to transmit coordinates directly to the FCS.

This paper focuses on describing the operational environment, architecture, and functioning of a custom-developed ballistic software module intended for a FCS. Although the module is designed with particular attention to contemporary trends and the specific requirements of mortar fire with minimal modifications, adaptation and application across a broader spectrum of artillery systems, thereby achieving a high degree of universality. The architecture presented in this work builds upon established principles of ballistic computation while introducing a practical, platform-independent implementation that balances computational efficiency with operational flexibility.

Corresponding author: M.Marković;
E-mail: mdmarkovic@mas.bg.ac.rs
DOI: 10.5937/OTEH26016M

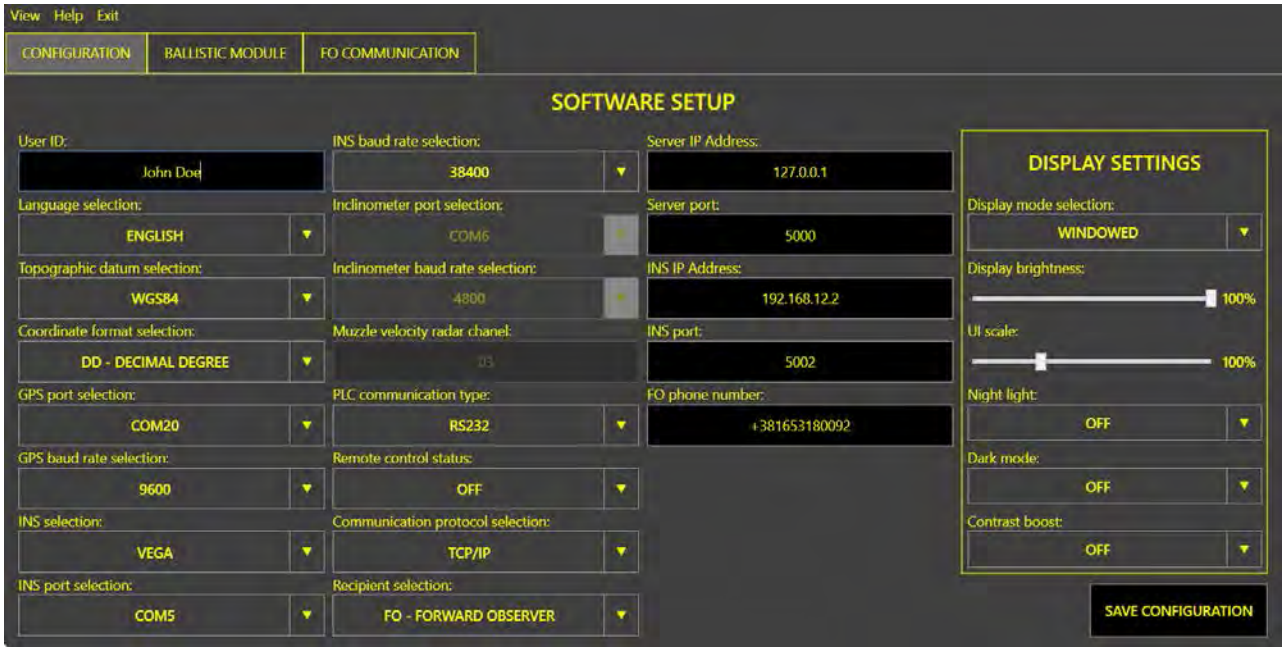


Figure 1. Configuration - Software Setup (example)

2. BALLISTIC MODULE FUNCTIONAL OVERVIEW

The process of calculating firing elements is implemented through an integrated ballistic computer algorithm, which begins with system initialization and the input of initial conditions such as firing position and target coordinates, internal ballistic parameters for the selected weapon, meteorological conditions, and environmental factors. The system then proceeds to solve the set of six-degree-of-freedom (6-DOF) differential equations of external ballistics, while optionally applying advanced correction models such as the Modified Point Mass Model (MPMMT) for enhanced computational efficiency in longer-range or rapid-response scenarios.

After each simulation, the system verifies whether the required range has been achieved by comparing the desired and actual impact point against the effective destruction radius of the munition. If a deviation is detected, the algorithm iteratively adjusts the initial condition vector primarily elevation and azimuth using predefined step sizes or optimization methods, and repeats the simulation until convergence is achieved. Once the target criteria are met, final refinements are applied, accounting for dynamic factors such as wind, projectile drift, and, if applicable, specialized engagement modes like Multiple Rounds Simultaneous Impact (MRSI) or linear target patterns.

In the concluding phase, the system outputs the complete firing solution, including quadrant elevation, azimuth, muzzle velocity, and any sequence-specific parameters, while simultaneously logging all computational steps and enabling real-time data exchange with higher-level command, control, and communication systems for seamless integration into the broader fire control architecture. A more detailed derivation of the computational algorithms is available in [1,3,6,13].

3. BALLISTIC MODULE ARCHITECTURE

The software environment of the ballistic module consists of three main tabs: Configuration, Ballistic Module, and Communication, as shown in Figure 1. Within the Ballistic Module section, there are two windows: Fire Preparation and Fire Elements Calculation, depicted in Figures 2 and 3, respectively, while Figure 4 presents the Communication tab.

3.1 Configuration – Software setup

The Configuration tab (Figure 1) enables initial setup of the FCS, allowing adaptation to specific operational scenarios, hardware configurations, and network architectures.

User Identification and Auxiliary Fields – The *User ID* field allows entry of a unique operator identifier (name, tactical designation, or internal code) used in network communications. The *FO Phone Number* field is used primarily during polygon testing and development phases for standard telephone communication with the forward observer.

System Configuration – Dropdown menus enable selection of language, topographic datum, and coordinate display format. Communication ports and baud rates are configurable for GPS, INS, and inclinometer devices. Additional selections include the inertial navigation system type (LYRA or VEGA), muzzle velocity radar channel, PLC communication type (RS232 or ETHERNET), and remote control status. For network

communication, the protocol (TCP or UDP), recipient, server IP address and port, as well as INS IP address and port, are defined. The *Save Configuration* button permanently stores all settings, enabling pre-adaptation to the platform and environment, thereby increasing efficiency and reducing firing data calculation time.

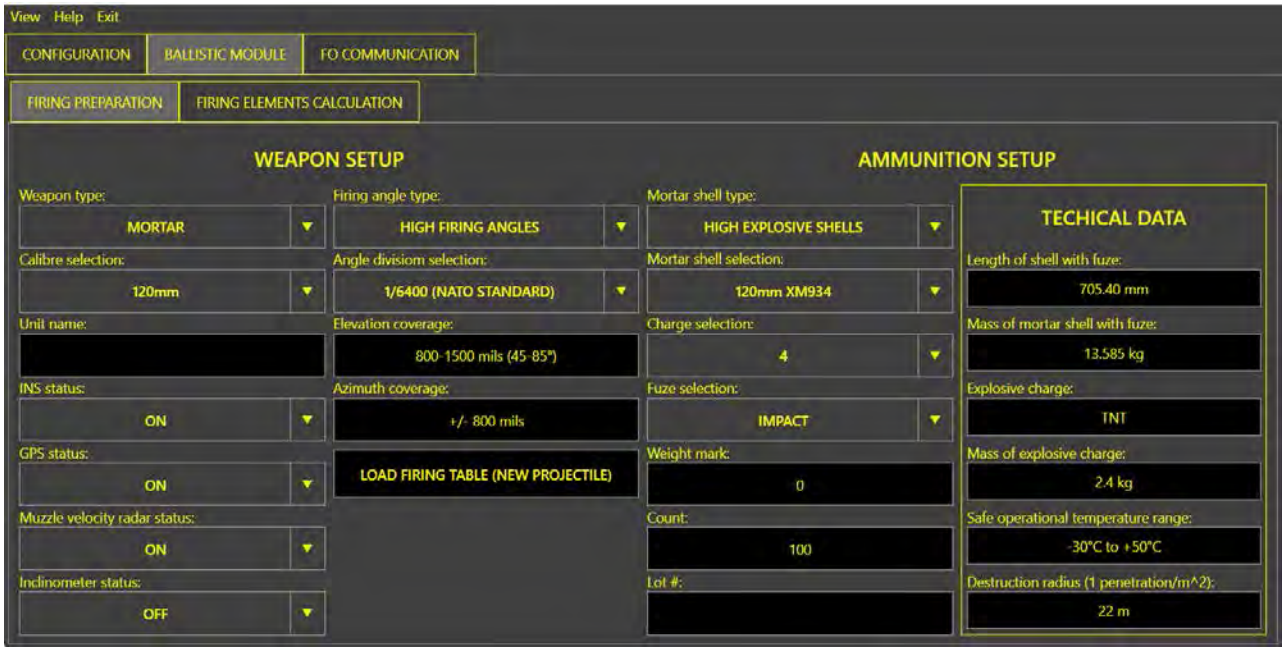


Figure 2. Ballistic Module - Firing Preparation (example)

3.2 Ballistic Module – Firing Preparation

The Ballistic Module tab, Figure 2, is the central computational component of the fire control system. It enables complete preparation and calculation of firing elements based on weapon type, ammunition type, firing position and target characteristics, and environmental conditions. The tab is divided into two logical sections: *Firing Preparation* and *Firing Elements Calculation*.

Firing Preparation – This section allows preparation (selection of weapon and ammunition) prior to calculating firing elements, enabling the software to support multiple weapon types and calibers.

Weapon and Ammunition Selection – The *Weapon Type* dropdown enables selection of the artillery system (howitzer, multiple rocket launcher, or mortar). The *Calibre Selection* dropdown is dynamically linked, displaying available calibers based on the selected weapon: Mortar (60 mm, 81 mm, 82 mm, 120 mm), Howitzer (155 mm), and MLRS (107 mm, 122 mm, 128 mm). The *Mortar Shell Type* dropdown (or *Projectile Type* / *Rocket Type*, depending on weapon) selects the ammunition category (HE, Smoke, Illuminating). The *Mortar Shell Selection* dropdown filters compatible ammunition based on the combination of weapon type, caliber, and ammunition category. The *Charge Selection* dropdown displays only propellant charges compatible with the selected ammunition (disabled for MLRS). The *Fuze Selection* dropdown allows selection of the fuze type (e.g., Impact, Super quick, Delay action, UT M39 P1, MRV-U). The *Weight Mark* field is used for ammunition weight designation.

Technical Data – This section provides a non-interactive informational display of the currently selected ammunition's technical characteristics, including length, weight, explosive weight and type, temperature range, and destruction radius.

System Status and Operational Fields – The *INS Status*, *GPS Status*, *MVR Status*, and *Inclinometer Status* dropdowns enable or disable the use of these sensors. The *Firing Angle Type* dropdown selects the firing regime (High Firing Angles or Low Firing Angles), which directly affects the elevation coverage display. The *Angle Division Selection* dropdown sets the angular conversion standard (360° = 6400 mils for NATO or 360° = 6000 mils for Russian standard). The *Azimuth Coverage* field displays the permissible traverse limits for the weapon. The *Unit Name* field allows entry of the organizational unit (e.g., battery, platoon), which can be combined with the previously defined *User ID* for network identification. The *Count* field specifies the number of available rounds, and the *Lot #* field records the production lot number for logistical tracking.

Load Firing Table – The *Load Firing Table* button allows uploading precomputed firing tables (.csv, .txt, .xlsx, or .xls format) for projectiles that are not part of the system's standard database and for which mass-inertial and aerodynamic data are unavailable for numerical trajectory simulation.

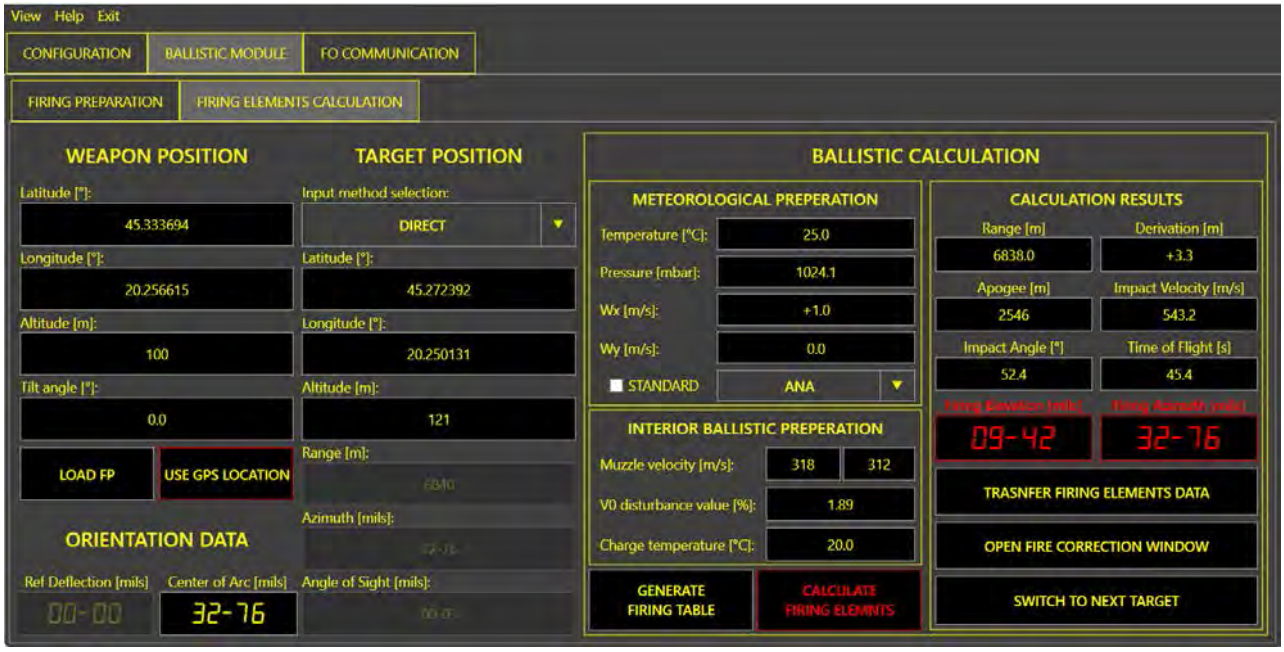


Figure 3. Ballistic Module - Firing Elements Calculation (example)

3.3 Ballistic Module – Firing Elements Calculation

The Firing Elements Calculation tab, shown in Figure 3, represents the core computational section of the ballistic module, where elevation and azimuth angles are determined based on predefined weapon and ammunition parameters combined with inputs entered within this section.

Weapon Position – This section defines the geographic location of the firing position, including latitude, longitude, altitude, and platform tilt angle. Accuracy of these inputs directly affects ballistic precision. Three input methods are available: manual entry, loading from a preconfigured firing position file (*Load FP* button), or reading from a GPS device (*Load GPS Location* button). Platform tilt angle can be entered manually or read from an inclinometer. Input formats depend on the topographic datum and coordinate format defined in the Software Setup tab.

Target Position – This section defines the geographic location of the target. Input accuracy directly impacts ballistic calculation precision. The *Input Method Selection* dropdown offers two modes:

- *Direct* – manual entry of latitude, longitude, and altitude.
- *Indirect* – target is defined relative to the firing position by entering range, azimuth, and site angle; the software then calculates geographic coordinates.

When the forward observer transmits target coordinates via the communication channel, the software automatically populates the relevant target information.

Meteorological Preparation – This section defines surface meteorological conditions affecting external ballistics, including temperature [°C], pressure [mbar], longitudinal wind (wx) [m/s], and crosswind (wy) [m/s]. The *Standard Atmosphere* checkbox allows selection of predefined standard conditions:

- *ANA (Artillery Normal Atmosphere)*
- *ICAO (International Civil Aviation Organization)*.

Interior Ballistic Preparation – This section defines parameters affecting interior ballistic calculations. Muzzle velocity is displayed with the prescribed nominal value on the left. If a muzzle velocity radar is used, the measured value appears on the right, along with the calculated percentage deviation, which is used to correct the firing elements (elevation angle). Propellant grain temperature is also defined, as it directly affects muzzle velocity through empirical coefficients, and together with barrel wear determines the total velocity deviation for fire control corrections

Calculation Execution – The *Generate Firing Table* button initiates a series of iterative ballistic calculations over a defined elevation range, producing a complete firing table in .pdf or .xlsx format for the selected weapon and ammunition combination. The *Calculate Firing Elements* button performs the final numerical calculation, verifying that all required inputs (weapon, ammunition, positions, meteorological parameters) are defined before computing the firing elements.

Figure 4. Forward Observer Communication

Calculation Results – This section displays the key outputs necessary for executing the fire mission. The primary calculated firing elements are *Firing Elevation* and *Firing Azimuth*, which represent the final angles to be transmitted to the weapon's executive system. Additional trajectory parameters are also displayed, including achieved range, deflection, maximum ordinate, impact velocity, impact angle, and time of flight.

The following auxiliary functions are available:

- *Transfer Firing Elements Data* – sends the calculated firing elements to the weapon's executive mechanisms.
- *Open Fire Correction Window* – opens a window for correcting firing elements, showing deviations from tabular values and allowing manual adjustment of elevation and azimuth based on forward observer reports of impact points.
- *Switch to Next Target* – resets current results and transitions to a new target for subsequent fire missions.

3.4 Forward Observer Communication

The FO Communication tab serves, Figure 4, as the communication interface between the weapon system and the forward observer. This section enables establishment, monitoring, and management of network connections, as well as real-time data exchange between connected entities. It is primarily designed to automatically process target data and projectile impact point data when transmitted by the forward observer. This functionality requires a radio device at both the observer and weapon locations.

Communication Service Management – The *Start Server* button initiates the communication service on the weapon system, assigning it the server role in the client-server architecture. The system opens the IP address and port defined in the Software Setup tab, initializes communication sessions, and enables forward observers to connect. Upon successful startup, the *Network Status* indicator turns green. The *Close Server* button safely terminates the communication service when needed.

Recipient Group – This section displays and enables selective communication with connected forward observers. The system supports up to four simultaneously connected observers, identified as FO1, FO2, etc. When an observer connects, the corresponding button becomes active and its status indicator turns green. These buttons allow selective message transmission to a specific observer.

Messaging – The *Formatted Messages* dropdown contains a list of predefined, standardized messages for rapid transmission to one or more connected observers, representing common operational commands or status information. The *Add/Remove Message* button allows modification of this list. The *Outgoing Message* field is a text area where the user composes messages to be sent selectively (via recipient group buttons) or to all connected observers simultaneously (via the *Send* button). The *Clear* button deletes the content of this field. All sent and received messages, along with connection status information, are displayed in the central panel of the tab.

4. PERFORMANCE EVALUATION

To assess the operational viability of the proposed ballistic module, two distinct scenarios were evaluated.

The first scenario represents full system configuration from a default state, encompassing all parameters within

the *Software Setup* tab, from user identification to communication protocols and display settings. The measured configuration time was approximately 40 seconds. While this duration is acceptable for initial integration and pre-deployment preparation, it does not reflect realistic tactical conditions, where systems are expected to be mission-ready with only mission-specific adjustments.

The second scenario reflects a realistic operational use case with predefined weapon type, caliber, and ammunition. Upon receiving target information (manually or via forward observer communication), the system performed ballistic calculations, including numerical trajectory integration and computation of firing elements and auxiliary trajectory parameters. The measured time from final input confirmation to displayed results was approximately 5 seconds.

The achieved 5 second calculation time aligns with real-time fire control requirements, where delays exceeding 10–15 seconds may compromise mission effectiveness, particularly against moving or time-sensitive targets. In comparison with legacy manual procedures using printed firing tables, which typically require 30–60 seconds depending on operator proficiency, the presented architecture provides a significant reduction in engagement time. The 40-second full configuration time, while not operationally critical, reflects the system's comprehensive configurability, enabling deployment across different weapon platforms (howitzers, mortars, MLRS) and hardware configurations without code modification. Together, these metrics demonstrate that the proposed architecture balances operational responsiveness with the flexibility required for multi-platform adaptability.

All measurements were performed on a hardware platform featuring an Intel Elkhart Lake Atom x6413E processor (quad-core, 1.50 GHz base frequency, up to 3.00 GHz boost), 8 GB of RAM, and 128 GB of internal storage.

5. CONCLUSION

The proposed custom-developed ballistic module architecture delivers a platform-independent fire control solution integrating configuration, ballistic computation, and forward observer communication into a unified software environment. With a full system configuration time of 40 seconds and a firing element calculation time of approximately 5 seconds compared to 30–60 seconds for legacy manual procedures the system meets real-time operational requirements while eliminating manual data entry errors. Its modular design supports seamless adaptation across howitzers, mortars, and multiple rocket launchers without code modification. With minor modifications, the system can also be integrated into a broader command and information system, enabling centralized fire control at the battery or battalion level. This architecture thus provides a universal solution compatible with existing artillery systems, combining speed, accuracy, flexibility, and interoperability as a foundation for next-generation fire control. The presented performance metrics were obtained under nominal conditions with optimal hardware resources and without network communication overhead. Future evaluation will include end-to-end latency with radio communication, performance under degraded hardware, simultaneous multi-observer connections, and live-fire validation against actual impact point accuracy.

Acknowledgment

The results presented here are the result of research supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under Contract 451-03-34/2026-03/200105 of 05.02.2026.

References

[1] Baranowski, L., Majewski, P., Szymonik, J., 2020, "Explicit form of the 'modified point mass trajectory model' for the use in Fire Control Systems", *Bulletin of the Polish Academy of Sciences: Technical Sciences*, Vol. 68, No. 5, pp. 1165-117

[2] Gacek, J., Leśnik, G., 2016, "Theoretical and Experimental Testing of the Factors of Mortar Projectiles Trajectory for a New Fire Control System", *Problemy Mechatroniki*, Vol. 7, No. 3, pp. 59-70.

[3] NATO Standardization Agency, 2009, *STANAG 4355 (Edition 3) - The Modified Point Mass and Five Degrees of Freedom Trajectory Models*, NSA/0454(2009)-JAIS/4355, NATO Standardization Agency, Brussels.

[4] Baranowski, L., 2013, "Feasibility analysis of the modified point mass trajectory model for the need of ground artillery fire control systems", *Journal of Theoretical and Applied Mechanics*, Vol. 51, No. 3, pp. 511-522.

[5] Chusilp, P., Charubhun, W., Ridluan, A., 2011, "Developing Firing Table Software for Artillery Projectiles using Iterative Search and 6-DOF Trajectory Model", in *Proceedings of the 2nd International Conference on Instrumentation, Control and Automation (ICA 2011)*, Krabi, Thailand, pp. 1-7.

[6] Petrovic, D.R., 2024, "Prediction of Artillery Correction Elements by Software for Ballistics Module of Fire Control Systems", in *Proceedings of the 11th International Scientific Conference on Defensive Technologies (OTEH 2024)*, Tara, Serbia, 9-11 October 2024, pp. 196-201.

[7] Ortac, S.A., Durak, U., Kutluay, U., Kucuk, K., Candan, C., 2007, "NABK Based Next Generation Ballistic Table Toolkit", in *Proceedings of the 23rd International Symposium on Ballistics*, Tarragona, Spain, 16-20 April 2007, pp. 1-9.

[8] Sowa, A.J., 2002, "The NATO Armaments Ballistic Kernel - the 1st item of an emerging suite of software component items for fire control systems", in *Proceedings of the 7th International Artillery and Indirect Fire Symposium and Exhibition*, Parsippany, NJ, USA.

[9] Luković, B., Petrović, V., Mrdak, V., 2013, "A Ballistics Module as a Part of the Fire Control System",

Military Technical Courier, Vol. 61, No. 3, pp. 217-230.

[10] FENIC'S, n.d., *Artillery Fire Control System - Fire Enhancement, Information, Command, Control and Communication System*, Company Brochure, [n.p.].

[11] Dyckmans, G., Goor, R., Vancraeynest, K., 2000, "ARTYFIRE - A Belgian Field Artillery Fire Control Computer Program", in *European Forum on Ballistics of Projectiles*, Saint-Louis, France, April 11-14, 2000, pp. 211-215.

[12] Giraud, M., Fleck, V. (eds.), 2000, *European Forum on Ballistics of Projectiles*, Proceedings of the Conference, Saint-Louis, France, April 11-14, 2000, Institut Franco-Allemand de Saint-Louis (ISL).

[13] Ostojić, A., 2025, *Development of a Ballistic Computer for Mortar Weapons*, Master's Thesis, University of Belgrade, Faculty of Mechanical Engineering, Belgrade. (in Serbian).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EXPERIMENTAL STUDY OF FRACTURED PROPELLANT INFLUENCE ON THE IMPULSE ROCKET MOTOR SAFETY

KATARINA NESTOROVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, katarina.nestorovic993@gmail.com
Military Academy, University of Defence, Veljka Lukica Kurjaka, 11042, Belgrade, Serbia

SAŠA ŽIVKOVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, sasavite@yahoo.com
Military Academy, University of Defence, Veljka Lukica Kurjaka, 11042, Belgrade, Serbia

MARINA SIMOVIĆ - PAVLOVIĆ
Teleoptic Gyroscopes, Filipa Visnjica 31, 11080 Belgrade, Serbia, simovicmarina99@gmail.com
Faculty of Mechanical Engineering, University in Belgrade, Kraljice Marije 16, 11000, Belgrade, Serbia

VESELIN ŽIVANOVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, veselin.zivanovic.best@gmail.com

JELENA MOJSILOVIĆ
University of Defence, Veljka Lukica Kurjaka, 11042, Belgrade, Serbia, mojsilovic.jeca@gmail.com

DARKO JANKOVIĆ
Military Academy, University of Defence, Veljka Lukica Kurjaka, 11042, Belgrade, Serbia, darkoojanko@gmail.com

Abstract: *Impulse rocket motors are group of solid propellant rocket motors, which main characteristic is extremely short operating time. The application of these rocket motors is of particular importance for missile weapons as busters or control thrusters. The total operating time of such rocket motor is in order of ten milliseconds and that complex processes are highly non-stationary. In the anti-tank rockets the impulse motor role is to accelerate the rocket to a certain velocity in a very short time, during the rocket's motion through the launch tube, and to provide the necessary ballistic characteristics (dynamic stability, greater range, shorter flight time to the target and the less dispersion of hits). Reliability of this type of rocket motors is of crucial importance in terms of operation safety. This paper presents an experimental study of how fractured propellant charge affects rocket motor performances and safety.*

Keywords: *impulse rocket motor, propellant, fracture, propulsion, pressure.*

1. INTRODUCTION

Rocket missiles are launched by a rocket motor which is an integral part of the projectile. In order for a rocket to accelerate to a certain veocity and to be able to maintain that velocity, overcoming air drag and gravity, it is necessary to generate a force called thrust [1-11].

Thrust is the reaction force that occurs when a certain mass is ejected from the propulsion group. The separation of mass is most often realized through the outflow of gases (gaseous products of combustion) where the direction of the generated thrust force is opposite to the direction of outflow of gases [1-10].

In order to ensure the generation of gases, it is necessary to have a certain source of energy. Chemical energy contained in solid propellant is converted into thermal energy through a propulsion reaction which is accompanied by the release of energy. Propulsion is performed in the propulsion chamber where the chemical energy is transformed into heat, while the thermal energy is converted into kinetic energy through the rocket motor nozzle [1-10].

In the anti-tank weapons, all energy required to solve the ballistic problem is provided by the rocket motor propulsion. The projectile of such a weapon must have a basic impulse rocket motor, which achieves initial velocity at the muzzle. Rocket can also have an additional rocket propulsion system that is activated during flight on the trajectory and improves its ballistic characteristics [12].

This type of weapon is used for shooting at short ranges (200 m to 500 m) by launching from the rocket operator's shoulder, therefore the reliability and safety of such systems must be at a very high level.

As the most common form of propellant for man-portable anti-tank rocket launcher are multiple small diameter thin tubes with. Such a propellant geometry is the most favorable because of:

1. definition and reproducibility of manufacturing

Corresponding author: K. Nestorović;
E-mail: katarina.nestorovic993@gmail.com
DOI: 10.5937/OTEH26017N

technology,

2. combustion without significant slivers,
3. elegant solution of propellants integration inside the rocket motor chamber and
4. control benefits [1].

The basic internal – ballistics (IB) parameters of the rocket motor are maximum pressure, mean effective propulsion time and total impulse.

2. THEORETICAL BASIS

The rocket motor operating regime calculation can be conducted using IB coefficients that can be calculated theoretically, experimentally or semi-empirically (Figure 1) [1-2].

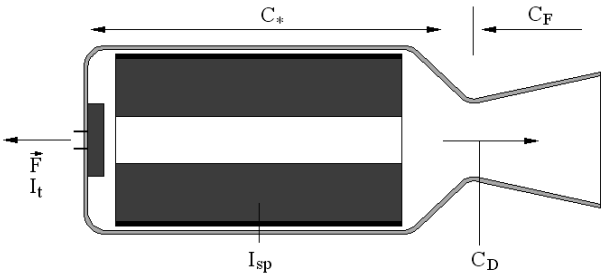


Figure 1. Basic model of propulsion product flow through the nozzle and characteristic process parameters [2].

Thrust of the rocket motor can be calculated based on the consideration of the motion of a body of variable mass (Mescherski's equation) and it is determined by the expression:

$$F = \dot{m}v_e + \left(p_e - p_a\right) A_e \tag{1}$$

where are:

$\dot{m}$ - mass flow rate,
v_e - nozzle exit velocity,
p_e – nozzle exit pressure,
p_a – atmospheric pressure and
A_e – nozzle exit area [1-2].

Nozzle exit pressure can be differentiated which depends on the difference (p_e - p_a):

- $p_e > p_a$ → under - expanded,
- $p_e = p_a$ → fully - expanded,
- $p_e < p_a$ → over – expanded [1-2].

Total impulse is the integral of thrust curve over the total operating time period τ and it can be calculated according to the equation:

$$I_t = \int_0^{\tau} F(t)dt \tag{2}$$

where is: τ – rocket motor operating time [1-2].

The specific impulse represents the thrust per unit propellants mass. Also, the specific impulse can be expressed as the ratio between the thrust coefficient c_F and the discharge coefficient c_D. It can be calculated according to the equation [1-2]:

$$I_{sp} = \frac{F}{\dot{m}} = \frac{I_t}{m} = \frac{C_F}{C_D} \tag{3}$$

The thrust coefficient represents efficiency of thrust generation, regarding nozzle flow efficiency and is dimensionless. It can be calculated according to the form:

$$c_F = \frac{F}{p_c A_t} \tag{4}$$

where are:
p_c – pressure of propulsion products,
A_t – specific nozzle throat area [1-2].

Discharge coefficient defines mass flow rate $\dot{m}$ trough throat area A_t, inducted with products pressure energy in chamber. It can be calculated according to the form:

$$c_D = \frac{\dot{m}}{p_c A_t} = \sqrt{\frac{\kappa}{RT_c}\left(\frac{2}{\kappa+1}\right)^{\frac{\kappa+1}{\kappa-1}}} \tag{5}$$

where are:
κ – heat capacity ratio,
R – gas constant,
T_c – chamber temperature [1-2].

The characteristic flow rate is a function of the propellant characteristics and propulsion chamber design, and it is independent of the nozzle characteristics. It can be calculated according to the equation [1-2]:

$$c^* = \frac{p_c \cdot A_t}{\dot{m}} = \frac{1}{c_D} \tag{6}$$

The output velocity of the propulsion product is calculated according to the equation [1-2]:

$$v_e = \sqrt{\frac{2\kappa}{\kappa-1} RT_c\left(1 - \bar{p}_e^{\frac{\kappa-1}{\kappa}}\right)} \tag{7}$$

Propulsion of the solid propellants takes place on the propulsion surface. Propulsion rate r is defined as a distance that crosses the flame front perpendicular to the free charging surface in a unit of time.

Pressure has a dominant influence on the value of propulsion rates and this dependence is expressed in form of Saint – Robert's law:

$$r = bp^n \tag{8}$$

where are: b – initial charge temperature function and n – the propulsion index which is a characteristic of rocket propellant and is always less than 1 [1-2].

Only in the case of propellant with frontal propulsion, the discharge of gases is carried out normally on the propulsion surface. In other cases, the gases move parallel to the propulsion surface, which affects the propulsion rate, which increases with increasing gas velocity along

the propulsion surface. This phenomenon is called erosive propulsion [1-2] [13].

The geometry of the nozzle (flow sections, number of nozzles and their position) is defined within the ballistic calculation of the rocket motor. During the flow of gaseous products through the nozzle, erosion of the nozzle neck can occur, which causes a reduction in the total impulse. Due to the expansion of the nozzle neck caused by erosion, the pressure drops and the operation of the rocket motor is prolonged. The erosion process occurs when the surface temperature of the nozzle neck reaches the melting temperature of the nozzle material. The most intense heat transfer from the propulsion surface to the flow cross-sectional surface is in the nozzle throat, so the melting temperature is reached closest there [1-2] [13].

The degree of expansion *(ε)* of the nozzle is a constant geometric characteristic that depends on the ratio of specific heats of the propellant and the ratio of pressures in the chamber and at the outlet of the nozzle. The calculation of the nozzle strives for the complete expansion of gases, with the aim of achieving the highest energy utilization. This is achieved if the pressure at the nozzle outlet is equal to the ambient pressure. The degree of expansion of the nozzle is determined by the equation [1-2]:

$$\varepsilon = \frac{A_e}{A_t} = \frac{\left(\frac{2}{\kappa+1}\right)^{\frac{1}{\kappa-1}} \cdot \left(\frac{p_c}{p_e}\right)^{\frac{1}{\kappa}}}{\left\{\frac{\kappa+1}{\kappa-1} \cdot \left[1-\left(\frac{p_c}{p_e}\right)^{\frac{1-\kappa}{\kappa}}\right]\right\}^{\frac{1}{2}}} \tag{9}$$

The chamber of the rocket motor is cylindrical shape, made of high-quality steel capable of withstanding the pressure of propulsion gases and their erosive action. The bottom of the rocket motor connects with a warhead over the thread. A stabilizer is set on the rear end of the rocket motor chamber.

3. EXPERIMENT

Since the man-portable anti-tank rocket is launched from the operators' shoulder, the rocket motor's operating time must be completed while the motor is in the tube. Such weapons are designed with a maximum rocket motor operation time of around $t_{max} \leq 18$ ms and the total impulse of around $I_t \approx 85$daNs [1]. For this type od weapons factor of safety must be maximum. The aim of this experiments was to show how fractured propellant influence on [5] [14]:

1. safety of the rocket operator,
2. performances of the impulse rocket motor, especially on pressure which is the most important rocket motor parameter in terms of safety.

The impulse rocket motor chamber is designed to have high mechanical properties. The high-strength aluminum alloy chambers usually can withstand pressure values of 1000 bar or more, which has been confirmed by experiments. The maximum pressure of the propulsion products in the rocket motor chamber is between 500 bar and 600 bar, and this is in case when propellant was previously exposed to extremely high temperatures (+50°C) for a period of time. In the case of ambient temperature (+20°C), the maximum pressure of combustion products is around 400 bar.

In this experiment two expected potentially critical failures are examined:

1. Due to the deterioration of the mechanical characteristics of the propellant and mechanical operational loads, different degree of fragmentation of certain number of tubes potentially may occur;
2. The bounded joint of the tubes to the grain holder may fail, which will lead to the ejection of certain number of the tubes through the nozzle.

For purposes of the experiments, two mentioned types of failure are simulated: some of propellant tubes were intentionally fractured or unbounded form the grain holder, in order to simulate propellant material deterioration, protentional damages of the rocket, bad exploitation conditions, etc. Since the propellant fractures cannot be easy detected, the aim of the experiment was to show whether such a flaw could cause a critical increase in pressure gradient [15] and the destruction of the rocket motor chamber, endangering the operator's safety.

Before realization of this experiment, propellant and the pyrotechnic mixture were tested by the method of vacuum stability and compatibility, described elsewhere [16].

The static tests of experimental rocket motors were realized on the test bench. First set of experiments were carried out at ambient temperature of +20°C, with intentionally fractured propellant. Propellant tubes are shredded to pieces 5×10 mm, and are shown in Figure 2. Shredded pieces of tubes are freely poured into chamber.

Figure 2. Intentionally fractured propellants tubes.

Second set of experiments are conducted with propellant grains with sertain percent of not bounded tubes, which are freely placed into the chamber as free-standing grains.

All static tests were carried out in experimental rocket motors, which free volume fully corresponds to a flight rocket motor. Changes in the pressure of the combustion products inside the chamber was measured during the

rocket motor operating time.

Experiment 1. Tube fracture.

In the first experiment, 4 of 37 or around 10% of the total number of propellant tubes were fractured. With such damage to the propellant, it was shown that the maximum pressure of the combustion products reaches over 350 bar, while operating time of the rocket motor is several milliseconds shorter than the projected operating time. Figure 3 shows diagram of the motor chamber pressure-time function.

Experiment 2. Tube fracture.

In the second experiment, 10 of 37 or around 27% of the total number of propellant tubes were fractured. Maximum pressure of the combustion products reaches 420 bar, while operating time of the rocket motor is similar like in the first experiment.

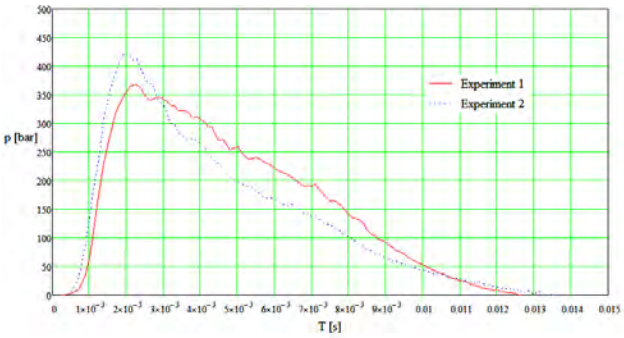


Figure 3. The rocket motor chamber pressure-time curves for experiments 1 and 2 with fracture tube.

The pressure-time curves are significantly changed compared to regular tests. The maximum pressure is increased, the pressure integral is reduced due to the ejection of unburned pieces of propellant, and the operating time is extended due to the pressure drop in the last stage.

Based on these two experiments, it can be concluded that a very significant fragmentation of the propellant grain is required in order to achieve a greater increase in pressure in the chamber. This type of propellant grain anomaly can lead to the motor failure only with tubes fragmentation in smaller pieces and a larger percentage of tubes.

Experiment 3. Joint failure.

In the third experiment, 8 of 37 or around 20% of the total number of propellant tubes were unbounded – free-standing. The maximum pressure of the combustion products reaches 340 bar, while operating time of the rocket motor is extended including significant secondary pressure peek. In this motor configuration complex internal ballistics phenomena are present, including movement of free-standing tubes. Free standing tubes are partially trapped in initial configuration, standing between chamber walls and bounded tubs. During the propellant combustion, products flow cause ejecting force on the free-standing tubes and they can randomly leave the motor partially burned. During the tube's passing through the nozzle, throat choking can occur, which can be explanation for secondary peek in pressure diagram.

Experiment 4. Joint failure.

In this experiment, 16 of 37 or around 37% of the total number of propellants tubes were unbounded. The maximum pressure of the combustion products reaches 380 bar, while operating time of the rocket motor is stilly extended. In this case two smaller secondary pressure peeks can be noticed. In this configuration free standing tubes are not rigidly confined like in previous experiment, so probably throat choking is less distinct.

Experiment 5. Joint failure.

In the last experiment, 22 of 37 or around 60% of the total number of propellants tubes were unbounded. The maximum pressure of the combustion products reaches 400 bar, while operating time of the rocket motor is also stilly extended, without secondary pressure peeks.

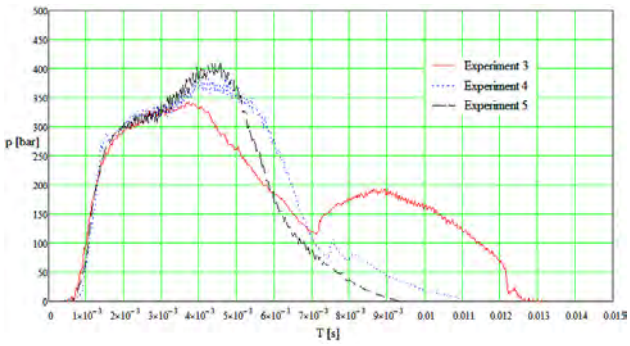


Figure 4. The rocket motor chamber pressure-time curves for experiments 3, 4 and 5.

The pressure-time curves are significantly changed in comparison to regular tests. The maximum pressure is increased, the pressure integral is reduced due to the unburned grains ejection. Assuming that every unburned grain ejection is causing some level of throat choking, configuration, type and percentage of ejecting tubes cause pressure increase which differ pressure diagram.

5. CONCLUSION

In this paper, an experimental study of fractured propellant influence on the impulse rocket motor safety was performed. From the IB parameters of the impulse rocket motor, the emphasis is placed on the maximum pressure of propulsion products inside the chamber as the most important safety parameter of the motor.

Five static tests were carried out with different amounts of fractured propellant tubes in order to check the safety impact on the rocket operator, as well as to check the changes in the pressure gradient depending of the increase in the propulsion surface.

In all of the five experiments, pressure gradient of the propulsion products was not significantly increased and did not exceed 415 bar (even in the case of severe damage to the propellant tubes - up to 60%), which corresponds to the nominal maximum pressure of the propulsion

products.

The operating regime of the rocket motor was not correct, which affects the operating time and which would potentially affect the flight parameters of the rocket, but the requirements during the desing of the weapon guarantee the safety of the rocket operator even in case of major damage to the propellant.

Nowadays, modern variants of man-portable anti-tank rocket launchers are in use in current warfares. An experimental study of this type provides a clear insight into the importance of safety over functionality in weapon systems that are launched by a rocket operator, in cases of unforeseen failures.

Acknowledgments

This work was supported by the Ministry of Science, Technological Development and Innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] GLIGORIJEVIĆ N., DIVAC S., LEMIĆ D., PETRIĆ A., SAVKOVIĆ M., ČOLAKOVIĆ M., VUČUROVIĆ O., JOVANOVIĆ V.: *Rocket Motors with Solid Propellant,* Military Technical Institute Belgrade, Serbia, 2013, ISBN 978-86-81123- 63-8.

[2] FERFOURI A., JERKOVIĆ D., KARI A., ŽIVKOVIĆ S., KARIĆ M.: *Computational Analysis of Rocket - Assisted Projectile Trajectory*, Scientific Technical Review, 2024, Vol. 74, No. 2, pp. 9-16.

[3] GLIGORIJEVIĆ N., RODIĆ V., JEREMIĆ R., ŽIVKOVIĆ S., SUBOTIĆ S.: *Structural Analysis Procedure for a Case Bonded Solid Rocket Propellant Grain,* Scientific Technical Review, 2011, Vol.61, No.1.

[4] GLIGORIJEVIĆ N., ŽIVKOVIĆ S., SUBOTIĆ S., KOZOMARA S., NIKOLIĆ M., ČITAKOVIĆ S.: *Side Force Determination in the Rocket Motor Thrust Vector Control System,* Scientific Technical Review, 2013, Vol. 63, No. 1, pp. 27-38.

[5] GLIGORIJEVIĆ N., BOULAHLIB M.A., ŽIVKOVIĆ S., SUBOTIĆ S., KOZOMARA S., NIKOLIĆ M.: *Solving an Irregular Burning Problem in a Small Rocket Motor,* Scientific Technical Review, 2014, Vol. 64, No. 2, pp. 3-13.

[6] GLIGORIJEVIĆ N., ŽIVKOVIĆ S., SUBOTIĆ S., PAVKOVIĆ B., NIKOLIĆ M., KOZOMARA S., RODIĆ V.: *Mechanical Properties of HTPB Composite Propellants in the Initial Period of Service Life,* Scientific Technical Review, 2014, Vol. 64, No. 4, pp. 8-16.

[7] GLIGORIJEVIĆ N., ANTONOVIĆ S., ŽIVKOVIĆ S., PAVKOVIĆ B., RODIĆ V.: *Thermal and Acceleration Load Analysis of New 122 mm Rocket Propellant Grain,* Scientific Technical Review, 2016, Vol. 66, No. 3, pp. 3-11.

[8] GLIGORIJEVIĆ N., ŽIVKOVIĆ S., RODIĆ V., ANTONOVIĆ S., MILOJKOVIĆ A, PAVKOVIĆ B., NOVAKOVIĆ Z.: *Reliability of Solid Rocket Propellant Grain Under Simultaneous Action of Multiple Types of Loads,* Scientific Technical Review, 2016, Vol. 66, No. 4, pp. 8-14.

[9] TANASKOVSKI B., NEŠIĆ J., JELISAVAC LJ., BOGDANOV J.: *A Numerical Method for Solid Propellant Grain Design*, Scientific Technical Review, 2018, Vol. 68, No. 2, pp. 48-54.

[10] RADULOVIĆ L., ŽIVKOVIĆ S., ANTONOVIĆ S., GLIGORIJEVIĆ N: *Refurbishing Project of Rocket Motor for a Rocket Artillery Projectile,* Scientific Technical Review, 2024, Vol. 74, No. 1, pp. 11-20.

[11] NESTOROVIĆ K., SIMOVIĆ – PAVLOVIĆ M., PAVIĆ M., STOJKOVIĆ S., BOSILJČIĆ S.: *Probability of Hitting and Destroying a Surface Target for Artillery Rockets*, Scientific Technical Review, 2024, Vol. 74, No. 2, pp. 54-58.

[12] MARJANOVIĆ G., KOZOMARA S., STOJKOVIĆ S., GLIGORIJEVIĆ N., ŽIVKOVIĆ S., PAVKOVIĆ B.: *State of the Propulsive - Group Development for Enhanced Anti - Armored Guided Missile,* Scientific Technical Review, 2018, Vol. 68, No. 3, pp. 56-63.

[13] STOJKOVIĆ S., GLIGORIJEVIĆ N.: *Effect of Erosive Burning in Rocket Motor on the Ballistics of Rocket Flight*, 11th International Scientific Conference on Defensive Technologies OTEH, Serbia, Tara, 9-11 October 2024.

[14] GLIGORIJEVIĆ N., RODIĆ V., PAVKOVIĆ B., BOGDANOV J., ŽIVKOVIĆ S., BAJIĆ Z.: *Effect of Cyclic Loads on Structural Damage of Rocket Propellant Grain,* Scientific Technical Review, 2019, Vol. 69, No. 2, pp. 8-16.

[15] GLIGORIJEVIĆ N., NEŠIĆ J., BOGDANOV J., PAVKOVIĆ B., ŽIVKOVIĆ S., TANASKOVSKI B.: *Influence of the Ignition Mixture Particle Size on Rocket Motor Igniter Pressure Gradient,* Scientific Technical Review, 2018, Vol. 68, No. 1, pp. 33-39.

[16] KRSTOVIĆ M., STANČIĆ T., NESTOROVIĆ K., BAJIĆ D., BOGDANOV J., BAJIĆ Z.: *Vacuum Stability Evaluation of Pyrotechnic Mixture – Double Base Propellant Compatibility,* 28th Consecutive Seminar on New Trends in Research of Energetic Materials NTREM, Chezch Republic, Pardubice, 22-24 April 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTEGRATED TECHNICAL MODEL OF DETECTION AND RISK MANAGEMENT OF EXPLOSIVE REMNANTS OF WAR - CASE STUDU SMEDEREVO FORTRESS

ANITA KLIKOVAC
Ministry of Internal Affairs, Emergency management sector, Smederevo, Serbia, anitaklikovac@yahoo.com

SANELA VELJKOVIĆ
Vinča Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, Department of Physical Chemistry, Belgrade, Serbia, sanela.veljkovic@vin.bg.ac.rs

MARINA DABETIĆ
Vinča Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, Department of Physical Chemistry, Belgrade, Serbia, fmarina@vin.bg.ac.rs

Abstract: *The presence of explosive remnants of war represents a long-term security and technical challenge, especially in urban areas and zones of exceptional cultural and historical importance. The paper presents the development of an integrated technical model for detection and risk management of explosive remnants of war, applied to the case study of the Smederevo Fortress site. The aim of the research is to establish a systemic approach that combines geophysical detection methods, an analytical risk assessment model and operational response procedures, with the aim of reducing the probability of an incident and its consequences for the population, infrastructure and cultural assets. The risk assessment was developed through a matrix that includes the probability of presence and activation of explosive remnants of war, population density, proximity to infrastructure and terrain specifics. By integrating technical data with operational security procedures, a phased model of risk management was defined, which includes initial assessment, technical reconnaissance, perimeter security, intervention and post-operative evaluation. A special contribution of the work is reflected in the adaptation of technical methods to the specifics of the protected cultural site, where the limitations regarding excavation and intervention are expressed. The proposed model can be applied to other locations with similar historical and infrastructural characteristics, thereby contributing to the development of a standardized technical-security approach to managing the risk of explosive remnants of war.*

Keywords: *explosive remnants of war, risk assessment, Smederevo fortress*

1. INTRODUCTION

Explosive remnants of war (EOR) represent one of the most significant long-term security challenges in post-conflict societies. Although armed conflicts often ended decades ago, their material remnants continue to threaten human life, infrastructure, cultural property and the environment. The presence of unexploded explosive devices can limit the development of certain areas, make it difficult to carry out construction and infrastructure works and represent a constant source of risk for the local population and visitors. Contemporary research indicates that effective EOR risk management requires a systematic approach that combines technical detection, risk assessment and response planning to potential incidents (Novik et al., 2023). According to the Law on Disaster Risk Reduction and Emergency Management, explosive remnants of war are explosive devices and unexploded ordnance, with or without detonators, prepared for use or used, i.e. fired, launched, ejected, launched or placed, and did not perform their function, i.e. unexploded, abandoned ordnance that was left behind or thrown away and assets left behind after the explosion of the ordnance warehouse, i.e. in the process of ordnance production (Law on Disaster Risk Reduction and Emergency Management, 2018).

A special challenge is presented by cultural and historical sites that had a military function during their history or were exposed to war. In such locations, there is often the possibility of the presence of residual explosive devices, while at the same time there are significant limitations in the application of invasive research methods due to the need to preserve cultural heritage. Therefore, it is necessary to apply modern detection methods that enable the identification of potentially dangerous objects without endangering the integrity of the site (Barone, 2019).

Smederevo Fortress is one of the most important medieval fortifications in the Republic of Serbia. A special place in the history of the Smederevo Fortress is occupied by the explosion of the ammunition warehouse on June 5, 1941, which represents one of the greatest tragedies in the history of Serbia. In addition to the large number of human casualties and the destruction of the city, the explosion left long-term consequences for the area of the fortress, including the possibility of unexploded ordnance

Corresponding author: A. Klikovac;
E-mail: anitaklikovac@yahoo.com
DOI: 10.5937/OTEH26018K

remaining in certain parts of the site. The historical circumstances of this event are documented in detail in the monograph by Miroslav P. Lazić, which represents a significant source for understanding the origin and development of security risks in the area of the Smederevo Fortress (Lazić, 2021). The historical circumstances of this locality point to the need for continuous consideration of potential security risks associated with the possible presence of explosive remnants of war.

The development of geophysical methods has significantly improved the possibilities of detecting explosive devices. Magnetometric, electromagnetic and other non-invasive surveys enable the identification of underground anomalies and potentially dangerous objects with minimal disturbance of the research area. The special importance of these methods is reflected in their application on cultural heritage sites, where preserving the authenticity of the space is one of the basic requirements (Ibraheem et al., 2021). In addition to technical detection, modern approaches emphasize the importance of spatial analysis and visualization of risk areas, which improves the process of decision-making and planning of protection measures (Lacroix et al., 2013).

Based on the above facts, the subject of this paper is the development of an integrated technical model of detection and risk of explosive remnants of war on the example of the Smederevo fortress. The aim of the work is to establish a model that combines technical methods of detection, risk assessment and operational procedures of risk management in order to increase the safety of people, infrastructure and cultural heritage. The scientific contribution of the work is reflected in the proposal of an integrated approach that can be applied to other cultural and historical sites with similar characteristics.

2. THEORETICAL AND CONCEPTUAL FRAMEWORK

Explosive remnants of war represent a residual explosive community that was not activated during use or was left abandoned after the end of armed conflicts. Their presence is a long-term security problem because they can remain active for decades after hostilities have ended. The risk of EOR is reflected in the possibility of causing human casualties, material damage and restrictions on the use of space for infrastructural, economic and social development (Novik et al., 2023). The presence of EOR is not only a technical problem of removing hazardous materials, but requires a systemic approach that includes hazard identification, risk assessment, prioritization of actions and implementation of appropriate protection measures. Modern approaches to risk management are based on the integration of technical, organizational and security measures in order to reduce the probability of an incident and mitigate its consequences (Novik et al., 2023). Technical detection is the basis for identifying and locating explosive remnants of war. The development of geophysical methods made it possible to significantly improve the field research process, especially in conditions where the application of invasive methods is not possible or desirable. One of the most commonly used methods is magnetometry, which is based on registering anomalies in the Earth's magnetic field caused by the presence of metal objects in the underground. This method enables a quick search of large areas and efficient identification of potential locations where explosive devices may be located (Butler, 2003). In addition to magnetometry, electromagnetic methods that enable more precise determination of the position and characteristics of underground metal objects are also widely used. The combined application of several geophysical methods contributes to increasing the reliability of results and reducing the possibility of misinterpretation of data (Pašteka et al., 2021). The use of georadar (Ground Penetrating Radar - GPR) is particularly important for the purposes of researching cultural and historical sites. This method enables non-invasive investigation of underground structures and anomaly identification without damaging archaeological layers and building structures. This is why georadar is often used in research on cultural heritage and protected sites (Martinho & Dionisio, 2014). Modern research indicates that the greatest efficiency is provided by the integration of several geophysical methods with the application of geographic information systems (GIS), which enable the processing, visualization and spatial analysis of collected data. Such an approach contributes to more accurate mapping of risk zones and more efficient planning of detection and remediation activities (Martorana et al., 2023).

Risk assessment is a key element of the security management system in areas suspected of containing explosive remnants of war. In a broader sense, risk can be viewed as a function of the probability of an unwanted event occurring and the consequences that event may produce. In EOR, risk assessment includes the identification of potential hazards, the analysis of the probability of their activation and the assessment of possible consequences for the population, infrastructure, environment and cultural assets. Based on the results of the assessment, action priorities and appropriate protection measures are defined (Novik et al., 2023).

Risk management involves planning and implementing organizational, technical and operational activities aimed at reducing the probability of an incident occurring or mitigating its consequences. This process includes continuous data collection, conducting technical research, establishing security procedures and making risk assessment decisions. With cultural and historical sites, risk management has additional complexity due to the need to preserve cultural heritage. Therefore, modern approaches are based on the integration of technical detection methods, spatial analysis and operational planning into a unique risk management system. Such an integrated approach enables more efficient use of available resources and a higher level of protection of people and cultural assets.

3. RESEARCH METHODOLOGY

The research is based on the case study method, which is

one of the most frequently applied qualitative research methods in social and technical sciences. This method enables a detailed study of a specific phenomenon in its real environment, while considering all the factors that influence its creation and development (Yin, 2018). A case study is particularly suitable for researching security problems, as it allows the connection of theoretical knowledge with empirical data collected in a specific location (Stake, 1995).

The Smederevo Fortress, one of the most important cultural and historical sites in the Republic of Serbia, was chosen as the subject of research. The choice of this locality is based on its historical significance, military function during different historical periods, as well as on the fact that in its area in previous years there have been cases of finding explosive remnants of war that required the intervention of competent services. The application of the case study enables the analysis of the existing situation, the identification of security risks and the development of an integrated technical model of risk detection and management that can be applied to other cultural and historical sites with similar characteristics. Primary and secondary sources of data were used in the research. Secondary sources consist of domestic and international scientific and professional literature in the field of explosive remnants of war, geophysical detection methods, risk assessment, security management and cultural heritage protection. Primary sources include data on the recorded interventions of competent services in the area of the Smederevo Fortress from 1994 to the present, including information on found explosive devices, the circumstances of their discovery and implemented security measures. For the purposes of the research, 38 recorded interventions were analyzed, which represent the basis of the empirical part of the work.

The work uses a combined research approach that combines qualitative and descriptive research methods. The analysis of scientific and professional literature was used to review modern approaches to the detection of explosive remnants of war and risk management models. The analysis of the documentation enabled the systematization of data on recorded interventions at the site of the Smederevo Fortress. The descriptive method was applied to show the characteristics of the explosive devices found, the frequency of interventions and their spatial distribution, while the comparative analysis enabled the comparison of existing approaches to risk management with the proposed integrated model. By combining different research methods, a more comprehensive overview of the investigated problem is achieved and the reliability of the conclusions drawn increases (Creswell, 2018).

The analytical framework is based on the assessment of the risk of explosive remnants of war using a qualitative analysis of the available data. Risk assessment includes hazard identification, analysis of the possibility of their occurrence and assessment of possible consequences for people, cultural heritage and infrastructure (Novik et al., 2023).

Within the scope of the research, the following were analyzed:

- the number of recorded interventions by competent services;
- types of explosives found;
- locations of finds within the fortress complex;
- the circumstances under which the funds were found;
- potential impact on the safety of visitors and employees i
- possibilities of applying modern technical detection methods.

Based on the results of the analysis, the elements of the integrated technical model of risk detection and management were defined, which represents the central contribution of this work

Also, we have to look at the limitations during the research. The limitations of the research refer to the availability of data on all historical interventions, as well as the fact that no direct geophysical surveys of the terrain were carried out within this work. Therefore, the analysis is based on the available documentation, recorded interventions of competent services and relevant scientific literature. Despite the mentioned limitations, the data used enable a reliable assessment of security challenges and the formulation of proposals for an integrated technical model of risk management

4. INTEGRATED TECHNICAL MODEL OF DETECTION AND RISK MANAGEMENT OF EXPLOSIVE REMNANTS OF WAR

Risk management is a set of measures and activities that are carried out with the aim of implementing disaster risk reduction policy as well as administrative operational and organizational skills and capacities for their implementation (Law on Disaster Risk Reduction and Emergency Management, 2018). Managing the risk of explosive remnants of war is a complex process that requires the integration of technical, organizational and security activities into a single system. Modern approaches indicate that individual measures, such as detection or removal of explosive devices, are not sufficient if they are not connected to systematic risk assessment and planning of preventive activities (Novik et al., 2023). Authors Nenad Komazec and Katarina Janković point out that integrated models of risk management enable more efficient planning of activities, more rational use of resources and improvement of safety during work with explosive remnants of war. Although their model was developed for EOR transport, the basic principles can also be applied to risk management at cultural and historical sites (Komazec & Janković, 2023). Starting from the mentioned approaches, in this work an integrated technical model was developed for risk assessment and management at the site of the Smederevo Fortress.

- The structure of the integrated model

The proposed model consists of six interconnected phases.

Phase 1 - Data collection and analysis

The first phase includes the collection of historical sources, archival documentation, data on previous interventions by the police and specialized teams, cartographic documentation and all available information that may indicate the presence of explosive remnants of war. At the Smederevo Fortress, historical data on the military name of the locality, the warehouse explosion in 1941, as well as data on subsequent discoveries of explosive devices are of particular importance.

Phase 2 - Technical detection

Based on the collected data, a technical survey is carried out using modern geophysical methods. The most important methods include:

- magnetometry;

- electromagnetic detection;

- georadar (GPR);

- GPS and GIS technologies.

Combining the mentioned methods increases the reliability of detection and reduces the possibility of misidentification of underground objects (Butler, 2003; Martinho & Dionisio, 2014).

Phase 3 - Risk assessment

After the identification of potential explosive devices, a risk assessment is carried out.

They are analyzed:

- type of explosive device;

- balance of assets;

- the depth of the find;

- activation probability;

- the frequency of the presence of people i

- cultural and historical value of the locality;

- possible consequences for people, infrastructure and cultural property.

As Vladimir M. Cvetković and colleagues point out, quality risk assessment is the basis for adopting adequate protection measures in the emergency management system (Cvetković et al., 2022).

Stage 4 - Making a decision

Based on the risk assessment, the procedure is determined. They may include more:

- continuous monitoring;

- additional geophysical research;

- restriction of access to the locality;

- engagement of specialized EOD teams;

- removal or destruction of an explosive device.

This approach enables decision-making based on risk assessment, and not exclusively on the assessment of an individual event (Komazec & Janković, 2023).

Phase 5 - Operational implementation

The operational phase implies the implementation of all planned activities by the competent services. Includes:

- location security;

- evacuation of the population if necessary;

- identification of funds;

- neutralization or transport;

- documenting the intervention.

All activities must be conducted in accordance with applicable safety procedures and standards.

Phase 6 – Evaluation and improvement of the system

The last phase of the model includes the analysis of the implemented intervention, updating the database and creating new risk maps. Based on experiences from previous interventions, procedures are improved, priority zones for future research are determined and preventive activities are planned. Such an approach enables the risk management system to be dynamic and to continuously adapt to new knowledge. The proposed model was applied to the case study of Smederevo Fortress. The starting point was historical data, documentation about the location and records of ten interventions by competent services due to the discovery of explosive remnants of war.

Based on the analysis of data collected from 1994 until today, it is possible to identify the zone of increased risk, determine priorities for technical research and propose preventive measures that contribute to the safety of visitors and the preservation of cultural property.

The advantage of the proposed model is reflected in the fact that it combines historical data, technical detection, risk assessment and operational decision-making in a single system that can be applied to other locations with similar characteristics.

5. CASE STUDY – SMEDEREVO FORTRESS

The construction of the Smedrevo fortress from 1428 to 1439 represented a unique construction undertaking, not only in terms of construction and organization, but also in cultural, historical and symbolic terms. Despot Đurađ Branković simultaneously built a strong military stronghold as the greatest achievement of medieval Serbian military architecture, but also a new capital city as a dynamic center of medieval Serbia. The undertaking stands out for its scale, short construction time and applied military-strategic concept

It was built in a period of dramatic historical events at the contact area between Europe and Asia during the decline

of the Byzantine Empire and the birth of the new Ottoman Empire on its ruins. The Smederevo fortress, created at the intersection of space, epochs and historical breaks, experienced its fall in 1459, ending the historical course of the medieval Serbian state, and beginning the epic life of Smederevo in the traditions, as a unifying symbol of the heroism and tragedy of the Serbian people. The life of the city of Smederevo continued in the following centuries within the framework of the Ottoman and Austrian states, and in 1805, during the restoration of Serbian statehood, as the first freed city and seat of the Governing Soviet, it became the thread that connects medieval and modern Serbia in a broad historical arc.

The 20th century brought catastrophic destruction to the Smederevo fortress in the world wars. First in the fall of 1915, during the second attack on Serbia, and then in the large explosion of stored military assets on June 5, 1941 (Lazić, 2021) and during the Allied bombings in 1944. These destructions left permanent consequences for the cultural property itself, but also for the entire city, which, in addition to great material suffering, suffered great human sacrifices. After the Second World War, extensive research and restoration of the Smedrevo fortress followed, and at the same time, its new evaluation as an important part of the national monumental heritage when it was designated as a cultural monument of exceptional importance. The process of proving its universal values as a basis for becoming part of the UNESCO world heritage is underway. The historical development of the Smederevo Fortress and its wartime sufferings represent an important context for understanding contemporary security risks at this location. During various infrastructural works, archaeological research and conservation interventions, a large number of unexploded ordnance was found. These facts confirm the importance of a systematic approach to risk detection and management in the area of cultural property.

The historical development of the Smederevo Fortress and the events of the Second World War represent an important context for understanding the contemporary security risks at this site. The explosion of the ammunition warehouse on June 5, 1941 caused large-scale destruction and left permanent consequences, which is why even today, during various works and interventions, unexploded ordnance is found. These facts confirm the importance of a systematic approach to risk detection and management in the area of cultural property (Lazić, 2021).

Source: Lazić, M. P. (2021). The explosion in Smederevo on June 5, 1941/80 years later. Museum in Smederevo.

The Smederevo fortress was a military facility until 1947, when the army handed it over to the city administration, and since then research, restoration, and its arrangement and infrastructural equipment have continued successively. After the Second World War, large damages as a result of war destruction were largely repaired, the craters caused by the big explosion directly near the main entrance and many smaller craters from the bombing were buried. During these works, the systematic removal of unexploded ordnance was carried out, as well as those encountered during the works, but for the period up to 1994, there is no reliable record of them.

The Regional Institute for the Protection of Cultural Monuments Smederevo keeps records from 1994 to 2026 for all unexploded hand grenades, mortar mines and artillery shells of various calibers, which were found in the area of the Smederevo Fortress and its immediate surroundings. A large amount of lighters, shrapnel, mining cables, slow-burning sticks, paraffin, rifle ammunition and other non-lethal war equipment, which was not precisely recorded, was also found. The largest amount of these findings was discovered in the period from 2019 to 2024, when extensive archaeological excavations were carried out in the Varoška Gate 1 zone (in table VK1), which was located directly in the direction of the main shock wave of the explosion on June 5, 1941. That is, on the rim of the crater that was created on that occasion, and which was filled with water for a long time after the Second World War, only to be buried in the early fifties.

12 hand grenades, 5 mortar shells and 13 artillery shells were found in the area of Varoška Gate 1, while 3 more hand grenades and 5 artillery shells were found in other locations. The largest grenade was found in 1998 (caliber 305mm) during infrastructure works in the wider area of Varoška kapija 1. All unexploded ordnance discovered were taken over by members of the emergency department, and since 2019, archaeological research has been carried out under pyrotechnic supervision.

During the analyzed period in the area of the Smederevo Fortress, starting from 1994 until today, 38 interventions by the competent services were recorded due to the discovery of explosive remnants of war, as shown in Table 1. The collected data indicate that the finds occur in

different parts of the complex, which confirms the need for continuous monitoring of the site and the application of preventive measures.

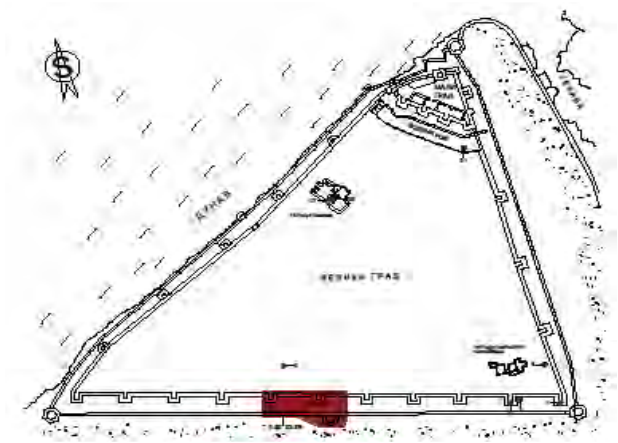
Table 1. Record of interventions

sn	year	Location	Type of EOR	Brief description of the event
1.	1994.	on tower 10	hand grenade	conservation works
2.	1998.	zone VK1	artillery shell	infrastrukturni radovi
3.	2011.	on tower 11	hand grenade	archaeological research
4.	2015.	along the Jezava rampart	artillery shell	conservation works
5.	2018.	on tower 11	hand grenade	conservation works
6.	2019.	zone VK1	hand grenade	archaeological research
7.	2019.	zone VK1	hand grenade	archaeological research
8.	2019.	zone VK1	hand grenade	archaeological research
9.	2019.	zona VK1	artillery shell	archaeological research
10.	2019.	zone VK1	artillery shell	archaeological research
11.	2019.	zone VK1	artillery shell	archaeological research
12.	2020.	zone VK1	artillery shell	archaeological research
13.	2020.	zone VK1	artillery shell	archaeological research
14.	2020.	zone VK1	artillery shell	archaeological research
15.	2020.	zona VK1	artillery shell	archaeological research
16.	2020.	zone VK1	hand grenade	archaeological research
17.	2020.	zone VK1	hand grenade	archaeological research
18.	2020.	zona VK1	mortar mine	archaeological research
19.	2020.	zone VK1	mortar mine	archaeological research
20.	2020.	zone VK1	mortar mine	archaeological research
21.	2020.	zona VK1	mortar mine	archaeological research
22.	2021.	zona VK1	artillery shell	archaeological research
23.	2021.	zone VK1	artillery shell	archaeological research
24.	2021.	zone VK1	artillery shell	archaeological research
25.	2021.	zone VK1	hand grenade	archaeological research
26.	2021.	zone VK1	hand grenade	archaeological research
27.	2021.	zone VK1	hand grenade	archaeological research
28.	2021.	zone VK1	hand grenade	archaeological research
29.	2021.	zone VK1	hand grenade	archaeological research
30.	2021.	zone VK1	hand grenade	archaeological research
31.	2023.	in the canal	artillery shell	infrastructure
		in front of the Little Town		works
32.	2024.	zone VK1	hand grenade	archaeological research
33.	2024.	zone VK1	minobacačka mina	archaeological research
34.	2024.	zone VK1	artillery shell	archaeological research
35.	2024.	zone VK1	artillery shell	archaeological research
36.	2024.	zone VK1	artillery shell	archaeological research
37.	2025.	hammam	hand grenade	archaeological research
38.	2025.	hammam	hand grenade	archaeological research

Source: Overview of registered EORs in the period from 1994 to 2026 from the documentation of the Regional Institute for the Protection of Cultural Monuments Smederevo, 2026.

Based on the data, the frequency of finding explosive remnants of war in the area of the Smederevo Fortress can be observed. By analyzing the place of discovery, the type of means and the circumstances of their discovery, it is possible to identify the zone of increased risk and set priorities for future technical research.

Based on the analyzed interventions, it can be concluded that the Smederevo Fortress is a site where there is still a risk of finding explosive remnants of war. The frequency of interventions confirms the justification of the application of an integrated technical model of risk detection and management, especially when performing construction, archaeological and conservation works.



Source: Map of registered EORs in the period from 1994 to 2026 from the documentation of the Regional Institute for the Protection of Cultural Monuments Smederevo, 2026.

7. DISCUSSION

The results of the research indicate that the Smederevo Fortress is a cultural and historical site where there is a continuous security risk from the discovery of explosive remnants of war. The analysis of 38 recorded interventions confirms that the presence of explosive remnants of war is not only a historical problem, but a

current security challenge that requires a systemic approach. The proposed integrated technical model enables the unification of historical data, results of technical detection, risk assessment and operational decision-making into a single management system. Such an approach is in line with modern tendencies in the development of risk management systems, which emphasize the need for an interdisciplinary approach and evidence-based decision-making (Novik et al., 2023). The application of modern geophysical methods, such as magnetometry, electromagnetic detection and georadar, enables non-invasive exploration of space, which is of particular importance for cultural heritage sites (Martinho & Dionísio, 2014). At the same time, the analysis of the case study shows that technical detection must be complemented by quality risk assessment, organizational measures and coordination of all competent institutions.

A special contribution of this work is reflected in the development of an integrated model that can serve as a basis for improving the risk management system not only in the area of the Smederevo Fortress, but also in other cultural and historical sites in the Republic of Serbia and the region where there is a possibility of the presence of explosive remnants of war.

8. CONCLUSION

Explosive remnants of war pose a long-term security problem that can threaten human lives, cultural property and infrastructure even decades after armed conflicts have ended. Localities that had a military function during history or were exposed to war activities are of particular importance, among which the Smederevo Fortress stands out. Based on the analysis of scientific literature and a case study, it was determined that existing approaches to risk management can be improved through the application of an integrated technical model that combines data collection, technical detection, risk assessment, decision-making, operational procedures and continuous evaluation. The analysis of recorded interventions in the area of the Smederevo Fortress confirms that the risk of finding explosive remnants of war still exists, which is why it is necessary to carry out preventive activities continuously, especially when carrying out archaeological research, conservation works and infrastructural interventions.

The proposed model can contribute to the improvement of people's safety, preservation of cultural heritage and more efficient planning and management of the activities of competent services. It is also important as part of the nomination process for entering the Smederevo Fortress on the UNESCO World Heritage List. In the complex and demanding process of nomination of the Smederevo Fortress for the UNESCO World Heritage List, which takes place in an extremely competitive environment, returning to the models and values from the beginning of the journey could decisively contribute to its successful completion. This would open a completely new chapter of preservation of established value systems, necessary for our future uniqueness and duration (Radovanović, 2023). Its application is not limited exclusively to the Smederevo Fortress, but can represent the basis for risk management in other cultural and historical sites with similar characteristics.

Future research should be directed towards the development of GIS databases, the application of drones, artificial intelligence and advanced geophysical methods to improve the process of detection and risk assessment of explosive remnants of war.

References

[1] Barone, P. M. (2019). Bombed archaeology: Towards a precise identification and a safe management of WWII's dangerous unexploded bombs. Heritage, 2(4), 2704–2711. https://doi.org/10.3390/heritage2040167

[2] Butler, D. K. (2003). Implications of magnetic backgrounds for unexploded ordnance detection. Journal of Applied Geophysics, 54(1–2), 111–125. https://doi.org/10.1016/S0926-9851(03)00071-6

[3] Creswell, J.W. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publications.

[4] Ibraheem, I. M., Aladad, H., Alnaser, M. F., & Stephenson, R. (2021). IAS: A new novel phase-based filter for detection of unexploded ordnance. Remote Sensing, 13(21), Article 4345. https://doi.org/10.3390/rs13214345

[5] Komazec, N., & Janković, K. (2023). An integrated risk management model for transporting explosive remnants of war: A case study in the Republic of Serbia. Acadlore Transactions on Applied Mathematics and Statistics, 1(1), 44–50. https://doi.org/10.56578/atams010105

[6] Lacroix, P., Delhay, E., & Wolff, E. (2013). Methods for visualizing explosive remnants of war. Applied Geography, 41, 179–194. https://doi.org/10.1016/j.apgeog.2013.04.007

[7] Lazić, M. P. (2021). The explosion in Smederevo on June 5, 1941/80 years later. Museum in Smederevo.

[8] Martinho, E., & Dionísio, A. (2014). Main geophysical techniques used for non-destructive evaluation in cultural built heritage: A review. Journal of Geophysics and Engineering, 11(5), 053001. https://doi.org/10.1088/1742-2132/11/5/053001

[9] Martorana, R., Capizzi, P., Pisciotta, A., Scudero, S., & Bottari, C. (2023). An overview of geophysical techniques and their potential suitability for archaeological studies. Heritage, 6(3), 2886–2927. https://doi.org/10.3390/heritage6030154

[10] Milosavljević, S., & Radosavljević, I. (2019). Fundamentals of Methodology in Political Science. Official Gazette.

[11] Novik, G. P., Abrahamsen, E. B., & Sommer, M. (2023). Improving the decision-making basis by strengthening the risk assessments of unexploded ordnance and explosive remnants of war. Safety Science, 160, Article 106065. https://doi.org/10.1016/j.ssci.2023.106065

[12] Pašteka, R., Hajach, M., Brixová, B., Mikuska, J., & Stanley, J. (2021). Real magnetic stripping method in

unexploded ordnance detection and remediation: A case study from Rohožník military training range in SW Slovakia. Contributions to Geophysics and Geodesy, 51(3), 415–437. https://doi.org/10.31577/congeo.2021.51.3.5

[13] Radovanović, D. M. (2023). Smederevo Fortress from capital town to World Heritage. Thematic Proceedings of the International Scientific Conference Educational Programs — The Future of World Cultural and Natural Heritage: 50 Years /1972–2022/ of the Convention on the Protection of World Cultural and Natural Heritage, 192–213. https://doi.org/10.18485/arh_whc50.2023.ch9

[14] Republic of Serbia. (2018). Law on Disaster Risk Reduction and Emergency Management (Official Gazette of the Republic of Serbia, No. 87/2018).

[15] Stake, R. E. (1995). The art of case study research. Sage Publications.

[16] Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage Publications.

[17] Cvetković, V. M. (2022). Disaster protection and rescue tactics. Scientific and professional society for risk management in emergency situations.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE ANALYSIS OF THEORETICAL AND EXPERIMENTAL MUZZLE VELOCITIES FOR THE EVALUATION OF POSSIBLE BARREL WEAR FOR THREE 30 mm BARRELS

VUK GARČEVIĆ
Military Technical Institute, Belgrade, vuk.garcevic@mod.gov.rs

ANDRIJA VELJOVIĆ
Military Technical Institute, Belgrade, andrija.veljovic@mod.gov.rs

IGOR RADISAVLJEVIĆ
Military Technical Institute, Belgrade, igor.radisavljevic@mod.gov.rs

Abstract: *The case study presents a comparison of the calculated muzzle velocity using the well-known internal ballistics Drozdov analytical mathematical model results and the experimental muzzle velocities obtained using a Doppler radar for the 30 x 165 mm AP-T projectile. The experiment was conducted during a field testing using three 30 mm barrels at the center for testing military equipment and weapons NIKINCI with the objective to determine their barrel wear. Therefore, the comparison of the measured and calculated muzzle velocities of the projectiles are the crucial parameters of relevance whether the barrels are in usable condition or not, which is analyzed and presented.*

Keywords: *Internal ballistics, Drozdov method, Doppler radar, muzzle velocity, 30mm barrel.*

1. INTRODUCTION

The case study examined the comparation of experimentally obtained muzzle velocities of three 30 mm barrels with characteristic shown in Table 1, with mathematically calculated muzzle velocities obtained using the internal ballistics Drozdov analytical mathematical model and tactical technical characteristics of the manufacturer for Ammunition Fig. 1, Table 2 and gunpowder in Table 3.

Table 1. Technical characteristics of the 30 mm barrel

Barrel Characteristic	Value
Caliber, d [mm]	30
Barrel length, l [mm]	2400

Table 2. 30 x 165 mm, AP-T Medium caliber ammunition technical characteristics [2]

Ammunition Characteristic	Value
Caliber, d [mm]	30
Muzzle velocity, v_0 [m/s]	970 ± 15
Max pressure of powder gasses, p_{max} [MPa]	360
Round mass, m_r [g]	853
Projectile mass, m [g]	400 ± 6
Tracer burning time, t [s]	Min. 3.5
Operating temperature [°C]	-40 to +50

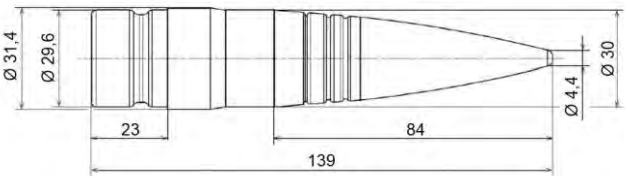


Figure 1. Projectile 30 x 165 mm [3]

The powder charge that was used for the 30 x 165 mm ammunition was the NC-48-7 perforated powder that is regulated by the standard SORS 8665 [10], the characteristics of the powder that are defined by the manufacturer [4].

Table 3. Technical data of the NC-48-7 powder charge [4]

Powder charge	Weight	Velocity	Pressure
NC-48-7	124 [g]	970 [m/s]	≤ 355 [MP_a]

2. MATHEMATICAL MODEL

To determine the projectiles muzzle velocity that is expected in the experiment, the muzzle velocity must firstly be calculated mathematically using classic theory of interior ballistics. For this case study the analytical model by Drozdov [1] was applied which is one of the classic theories of internal ballistics. Internal ballistics has the primary task of determining the relationship between the gunpowder pressure P [MPa] and the position of the projectile in the barrel. Also establishing the relation between the gunpowder pressure in the function of time respectively. Therefore, the Drozdov analytical model determines the internal ballistic parameters in the function of an independent variable such as time and displacement.

Corresponding author: V. Garčević;
E-mail: vuk.garcevic@mod.gov.rs

DOI: 10.5937/OTEH26019G

For this case study the parameter of analysis were the muzzle velocities. Therefore, it is the significant value that derives from the Drozdov analytical model. The ballistic property calculated using the Drozdov analytical model - maximum pressure in the barrel, is beneficial because it will also be compared with the projectiles manufacturing declaration [2].

The following are the fundamental equations that are used to solve the system of internal ballistics, determined by the Drozdov analytical method. [5].
Equation of energy balance:

$$pS_C\left(X_\psi + X\right) = f_b m_b \psi - \frac{k-1}{2}\varphi m \nu^2 \quad (1)$$

Equation of motion:

$$\varphi m \frac{d\nu}{dt} = pS_C \quad (2)$$

Projectile velocity:

$$\nu = \frac{dX}{dt} \quad (3)$$

Burning rate equation:

$$u_z = u_{z0} p \quad (4a)$$

with

$$\frac{dy}{dt} = \frac{u_{z0}}{r_0} p \quad (4b)$$

The burned propellant mass rate:

$$\psi = ky\left(1 + \lambda y + \mu y^2\right) \quad (5)$$

Grain surface burning rate:

$$\sigma = 1 + 2\lambda y + 3\mu y^2 \quad (6)$$

The approximate free volume length of the combustion chamber:

$$X_\psi = W_0 - \frac{m_b}{\rho_b}\left(1 - \psi\right) - \alpha m_b \psi \quad (7)$$

The essential values for determining the system of equations are grouped as:

- Barrel characteristics: S_C, W_0 and X_u,
- Projectile characteristics: m, p_0 and φ,
- Propellant characteristics: m_b, f_b, α, u_{z0}, ρ_b, k, $2r_0$, λ and μ.

The following assumptions are adopted based according to the Drozdov analytical method:

- The propellant grains burn according to the geometry hypothesis,
- The combustion takes place at medium pressure (Although in reality it changes from the bottom of the barrel to the bottom of the projectile),
- The combustion rate of the grain varies linearly with the pressure,
- Composition of the combustion products during the firing process does not change, so that the specific work and co-volume are constant, and
- During the firing process, the resistance of the air, the deformations of the exterior bullet surface and the interior walls of the barrel are neglected.

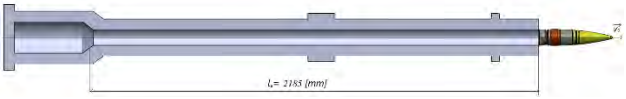


Figure 2. Graphic representation of the 30 mm grooved barrel length.

The results obtained using the Drozdov analytical model are presented in table 4 and are also presented in figures 3 and 4 respectively.

Table 4. Specific output values obtained using the Drozdov analytical method

Time [*ms*]	Projectile position [*m*]	Pressure [MP_a]	Velocity [*m/s*]
0.0	0.000	30.0	0
0.5	0.009	95.3	45.3
1.0	0.062	238.4	181.3
1.5	0.2094	346.9	430.5
1.73	0.3185	359.2	566.5
2.0	0.5052	227.0	694.3
2.5	0.8785	124.7	819.3
3.0	1.2891	80.5	892.0
3.5	1.7744	55.4	945.5
3.946	2.185	43.2	977.2

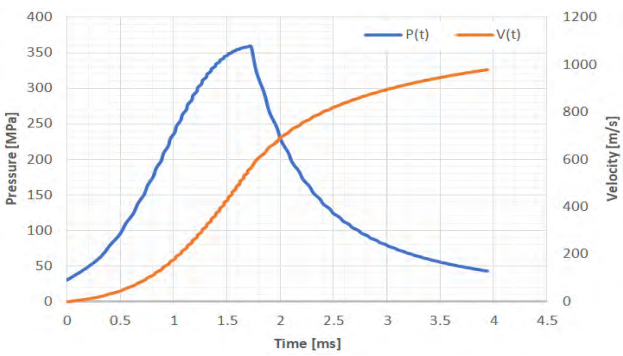


Figure 3. Pressure – Velocity – of the projectile as a function of time.

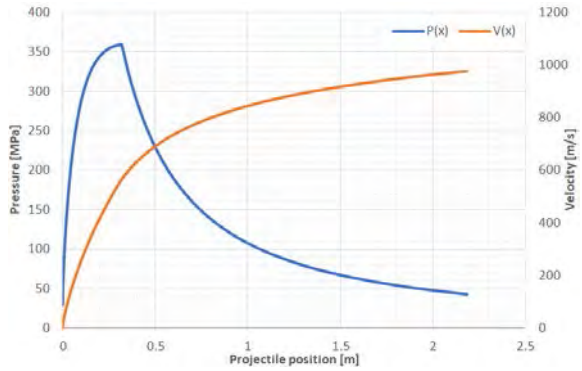


Figure 4. Pressure – Velocity – of the projectile as a function of the position of the projectile in the barrel.

The results obtained using the Drozdov analytical model show that the maximum pressure and muzzle velocity that can be expected are 359.2 MP_a and 977.2 m/s, respectively. The time that is needed for the projectile to leave the barrel is 3.946 ms.

3. FIELD TESTING

The field testing was conducted at the center for testing military equipment and weapons NIKINCI. Three 30 mm barrels were tested by measuring their muzzle velocities. The ammunition used to conduct the experiment was the 30 x 165 mm AP-T M90. The characteristics of the ammunition and the barrel are as shown in tables 1 and 2. Before the testing started, each barrel fired two rounds for conditioning.

The barrels were aligned at a horizontal elevation (0 deg). The first barrel fired eight rounds, second barrel fired thirteen rounds and the third barrel fired nine rounds. Each rounds muzzle velocity was measured using a Doppler radar that was placed adjacent to the muzzle of the barrel. This radar type works on the basis of the well-known Doppler effect [3], [4].

The testing was conducted in such a manner that the main objective was to measure the interior ballistic properties of the barrel without damaging the barrel itself. Since the only way to measure the pressure inside the barrel would be to use a piezoelectric pressure sensor or crusher-gauge [5],[6], by having to drill a hole in the side of the barrel and the cartridge of the ammunition round itself and therefore preventing further exploitation of the barrel. The resolution for this issue was to obtain the muzzle velocity of the projectile.
The obtained experimental results of the fired rounds for each barrel are shown in tables 5, 6 and 7 and graphically in figure 5, 6 and 7.

Table 5. Obtained experimental velocities of barrel no.1

Barrel no.1	
Round no.	Velocity [*m/s*]
1	955.0
2	957.0
3	956.2
4	949.4
5	952.0
6	953.3
7	944.3
8	953.4
Range (Max-Min)	12.7
Average velocity	952.6

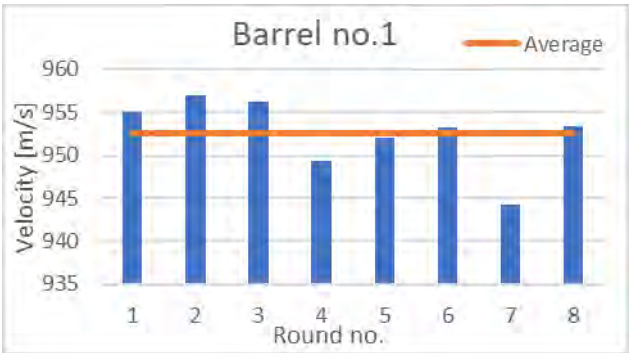


Figure 5. Graphical representation of the obtained muzzle velocities in the order of firing for barrel no. 1.

Table 6. Obtained experimental velocities of barrel no.2

Barrel no.2	
Round no.	Velocity [*m/s*]
1	953.8
2	952.6
3	954.9
4	960.0
5	954.4
6	946.0
7	966.2
8	964.2
9	953.3
10	949.1
11	953.5
12	953.8
13	951.4
Range (Max-Min)	20.2
Average velocity	954.9

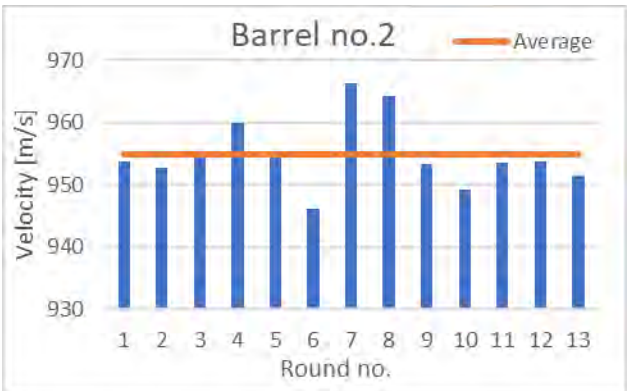


Figure 6. Graphical representation of the obtained muzzle velocities in the order of firing for barrel no. 2

Table 7. Obtained experimental velocities of barrel no.3

Barrel no.3	
Round no.	Velocity [*m/s*]
1	945.5
2	954.8
3	946.1
4	949.4
5	955.1
6	950.9
7	947.2
8	944.6
9	957.7
Range (Max-Min)	13.1
Average velocity	950.1

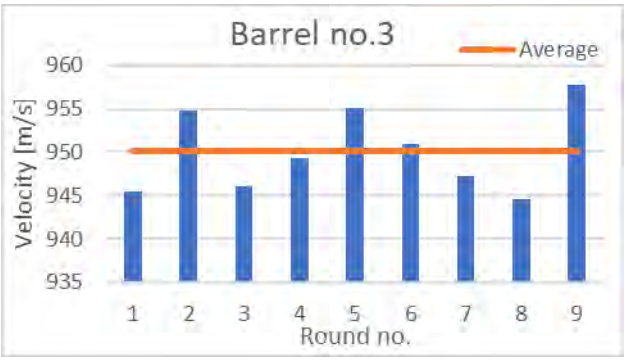


Figure 7. Graphical representation of the obtained muzzle velocities in the order of firing for barrel no. 3

4. RESULT COMPARISON

Firstly, a comparison between the manufacturer specified muzzle velocity defined in table no.2 and the muzzle velocity calculated in chapter 2 defined in table no.4 reveals a high degree of correlation, with negligible deviation from the nominal velocity.

Second, to compare the muzzle velocity with the values obtained through the mathematical model, the average experimental muzzle velocity for each barrel was utilized.

It is observed that for each barrel, the muzzle velocity falls below the lower limit defined by the manufacturer, as well as below the velocity calculated by the mathematical model. The differences between the analytical and experimentally obtained values range from 2.34% to 2.85%.

Given that these values are below the manufacturer's lower threshold, it is concluded that the three examined barrels do not meet the required criteria. Specifically, these barrels have entered the wear zone from an internal ballistics perspective.

Table 8. Difference of the analytical and experimental results for barrels no.1, no.2 and no.3

Average experimental results	Barrel no.1	Barrel no.2	Barrel no.3
Muzzle velocity [*m/s*]	952.6	954.9	950.1
Range max-min [*m/s*]	12.7	20.2	13.1
Difference of the analytical and experimental results. [%]	2.58	2.34	2.85

5. CONCLUSION

The comparison of the applied mathematical model obtained by the well-known method Drozdov for calculating internal ballistics parameters and experimentally gained muzzle velocities for three 30 mm barrels were performed in this case study. The field test was conducted using the 30 x 165 mm AP-T M90 ammunition. The objective of the experiment was to measure the interior ballistic properties of the barrel without damaging the barrel itself. Therefore, only the muzzle velocities were obtained using a Doppler radar.
Hence the obtained average experimental muzzle velocities for each barrel were compared to the analytical muzzle velocity.

The comparison concluded that each barrel muzzle velocity was below the analytically obtained muzzle velocity and the manufacturer's lower characteristic threshold.

The percentage difference for which the velocities are below the threshold that are defined by the analytical model are between 2.34% to 2.85%.

This concludes that from the internal ballistic aspect the barrels have entered the barrel wear zone. This conclusion could be better analyzed in future experiments using a piezoelectric pressure sensor or crusher-gauge. This would give an improved representation of the worthiness of the barrels from an internal ballistics aspect.

It would be useful to use a different type of known ammunition in further experiments to eliminate the possible suspicion of the used ammunition.

The results obtained in this case study encourage further experimental investigations in terms of internal ballistic parameters and barrel wear.

Acknowledgment

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03-34/2026-03/200325).

References

[1] BOUKERA ABACI,W., HRISTOV,N., ZIANE AHMED,N., JERKOVIC,D., SAVIC,S.: *Analysis of thermal and gas-dynamic characteristics of different type of propellant in small weapons*, Thermal Science: Year 2021, Vol.25, No. 6A, pp. 4295-4306,

[2] 30 mm x 165 AP-T for 2A42, 2A72, 2A38M eng..pdf.

[3] Пантовић ,П., Живковић, М., Миловановић, В., Милорадовић,Н.: Numerical analysis of the penetration process of a 30mm armour-piercing projectile, Војнотехнички гласник 2023, Vol. 71, issue 3, 2023.

[4] Baruti | Milan Blagojević - Namenska

[5] CVETKOVIC,M., Internal Ballistics, SŠOVID-VTA., Belgrade, Serbia, 1998.,

[6] TANCIC,LJ.: Practicum for *Internal Ballistics*, Odbrana Media Center, Belgrade, Serbia, 2012.

[7] TANCIC,LJ.: *Internal Ballistics Design*, Odbrana Media Center, Belgrade, Serbia, 2014.

[8] JARAMAZ,S., MICKOVIC,D.: *Internal Ballistics,* University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia 2011.

[9] VILICIC,J., GAJIC,M.: *Ballistics,* Military publishing house, Belgrade, 1979.

[10] Defense Standard of the Republic of Serbia – SORS 8665.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PERFORMANCE ANALYSIS OF SPECIAL SUBSONIC SMALL CALIBER WEAPONS FOR URBAN CONDITIONS

NEBOJŠA HRISTOV
Military Academy, University of Defence, Belgrade, nebojsahristov@gmail.com

DAMIR JERKOVIĆ
Military Academy, University of Defence, Belgrade, damir.jerkovic@va.mod.gov.rs

ALEKSA ANIČIĆ
Agency for Testing, Stamping and Marking of Weapons, Devices and Ammunition, Kragujevac, aleksa34@live.com

MILJKO PAVLOVIĆ
Company Milan Blagojevic – Namenska AD, Lucani, mpavlovic.mbn@gmail.com

IGOR EPLER
Military Academy, University of Defence, Belgrade, igorepler@gmail.com

Abstract: *The paper presents a comparative analysis of the ballistic performance of subsonic weapons of small caliber, which are suitable for use in urban conditions. The use of small-caliber subsonic weapons with shot suppressors is very effective for close range of 50 to 150 meters. The weapon suppressors enable operations that take place with reduced gunfire noise level and reduce the detection of the position of the shooters. Proper selection of ammunition systems is important for the effectiveness of subsonic weapons at close range. Automatic pistols, submachine guns and special subsonic rifles are used for such operations. The paper presents an analysis of significant ballistic parameters during the use of small-caliber firearms with silencers. The analysis includes a comparison of gunfire noise level, muzzle velocity and statistical indicators of precision and accuracy. The paper shows the experimental measurement of the gunfire noise level and the muzzle velocity. In addition to the reduction of gunfire noise level, the energy of the projectile that interacts with the target is also important. Based on the characteristics of the projectile, an estimate of the energy of the projectile during interaction with the target at different distances is presented.*

Keywords: *Muzzle Velocity, Sound Pressure Level, Energy of Projectile, Energy Density.*

1. INTRODUCTION

Special subsonic small caliber weapons for use in urban conditions are silenced personal weapons of small dimensions and mass intended primarily for tactical units. Standard rifle ammunition creates a localized gunshot noise during firing because of breaking the sound wall with the supersonic flight of the projectile on the trajectory, which easily detects the position of the shooter. By using subsonic projectiles with a weapon silencer, the sound of the gunshot is eliminated, which optimizes stealth and tactical advantage for operations at short distances in urban conditions. Among other things, weapons that use subsonic ammunition have lower recoil forces, which allows for more accurate hits, even in multiple bursts of fire. Weapon silencers (mounted or integrated) on this type of small arms, in addition to reducing sound, eliminate flash during firing, which increases the camouflage characteristics of tactical units. Among other things, short-barreled rifles (SBR) with folding stocks allow users to move freely in urban and difficult-to-access conditions, as well as facilitate the transportation of units during operations. Typical special subsonic ammunition is .300 Blackout (Figure 1a for SIG MCX Rattler Figure 2a), 9x19 mm (Figure 1b for Daewoo K7 Figure 2b), 9x39mm (Figure 1c for AS Val Figure 2c) and others similar. The mass of the projectile of the listed ammunition is generally in the range of 8 to 16 grams.

a) b) c)

Figure 1. Special subsonic ammunition: a) .300 Blackout, b) 9x19 mm, c) 9x39mm SP5

Generally, this type of special subsonic small caliber weapon is intended for effective action at distances of up to 200 meters, or rather, it provides the best results at distances of 50 to 100 meters.

Corresponding author: N. Hristov;
E-mail: nebojsahristov@gmail.com

DOI: 10.5937/OTEH26020H

a)

b)

c)

Figure 2. Special subsonic small caliber weapons: a) SIG MCX Rattler, b) Daewoo K7, c) VS VAL

The paper analyzes the performance efficiency of four representatives of weapons and ammunitions. The analysis considers the mass of the projectile, the muzzle velocity and the kinetic energy of the projectile in interaction with the obstacle at distances of 50 and 100 meters. The intensity of the gunshot sound is also analyzed as an important characteristic.

2. EXPERIMENTAL PERFORMANCE TESTING

Experimental tests were carried out on the automatic pistol M84 in 7.65x17 mm caliber, the HK MP5SD in 9x19 mm caliber, the VSS Vintorez in 9x39 mm caliber, and on the modified modular platform AM 20 in caliber 7.62/6.5x39 mm, for which a 9x39 mm caliber barrel was manufactured. During the experimental tests, the muzzle velocity of the projectile and the sound pressure level on the corresponding sample were recorded.

The sound level test of a firearm shot is complex due to the very short duration of the peak of the sound pressure level (SPL) function. The Bruel & Kjaer model 2250 shown in Figure 3 was used to register the SPL according to the MIL-STD-1474D testing standard. The device is equipped with software for sound level measurement, frequency analysis, and appropriate filters for measured signals (FFT analysis and vibration and frequency analysis). The sound level meter sensor is a 1/2" condenser microphone 4189, with a sensitivity of 50mV/Pa (corresponding to: 26dB equivalent to 1V/Pa) ± 1.5dB, a capacitance of 14pF (at 250Hz), which is connected to a ZC-0032 preamplifier with a nominal attenuation of 0.25dB [22].

Figure 3. Bruel & Kjaer model 2250 [22]

The MIL-STD-1474D specifies the position of the sound level meter, which is placed at a 90-degree angle, 100 centimeters to the left of the silencer tube outlet. Both the silencer and the sound level meter should be 160 centimeters above the short-cut grass. The SPL of the gunshot is recorded at this position, as shown in Figure 4. [20].

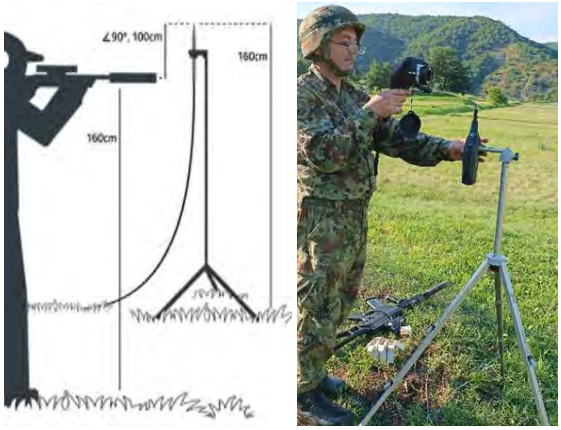


Figure 4. The noise meter position for measuring SPL [22]

To measure the muzzle velocity of the projectile, the GARMIN Xero C1 Pro Doppler ballistic chronograph shown in Figure 5, with the characteristics in Table 1, was used.

Figure 5. Chronograph GARMIN Xero C1 Pro

Table 1. Characteristics of Chronograph GARMIN Xero C1 Pro

Period	Production per period (pcs)
Measuring range	30 to 1600 m/s
Dimension	77 mm x 60.5 mm x 34.44 mm
Weight	105.6 g
Water rating	IPX7
Display type	Sunlight-visible, monochrome
Display size	133 mm x 44 mm, diagonal 55 mm
Battery type	Internal rechargeable lithium-ion battery; non replaceable
Battery life	Up to 2,000 shots or up to 6 hours
Operating range	-10° C to 55° C (14° F to 131° F)
Memory/history	50 sessions at max 100 shots per round before a Xero Core app sync is required to save data to phone
Interface	USB-C

During noise level measurements, for a detailed analysis of efficiency, in addition to *LCpeak* (the absolute maximum sound pressure level currently achieved, with a flatter frequency response that includes more low-frequency sounds), *LAeq* (the average sound level over the period, weighted to reflect typical human hearing) was also recorded.

As the first sample, the performance of the M84 automatic pistol, caliber 7.65x17 mm, with a silencer was measured (Figure 6).

The results of the parameter measurements are shown in Table 2.

Table 2. Measured performance for the automatic pistol M84, caliber 7.65x17 mm.

Group (10 shots)	Muzzle velocity (m/s)	*LAeq* (dB)	*LCpeak* (dB)
1	306,4	88	133,7
2	306,2	88,9	133,7
3	304,6	89,6	138,9
4	306	86,4	127,6
5	311,1	86,8	129,2
6	309,6	85,7	134,7
7	316,5	91,2	138
8	316,2	88,4	135
9	308	85,1	130,9
10	319,9	88,1	130,4
average	310,45	87,8	133,2
Min	304,6	85,1	127,6
max	319,9	91,2	138,9
Spread	15,3	6,1	11,3
Standard deviation	5,32	1,86	3,68

Figure 6. Automatic pistol 7.65 mm M84 with a silencer

In the second case, the characteristics measured for the submachine gun (SMG) HK MP5 SD, caliber 9x19 mm (Figure 7) are shown in Table 3.

Figure 7. SMG 9 mm HK MP5 SD

Table 3. Measured performance for the SMG 9×19 mm HK MP5 SD.

Group (10 shots)	Muzzle velocity (m/s)	*LAeq* (dB)	*LCpeak* (dB)
1	294,4	82,3	121,6
2	305,1	82,4	121,8
3	308,2	80,7	119,4
4	299,3	82,3	121,1
5	303,5	83,7	120
6	303,5	83,4	121,3
7	294,2	85,2	121,6
8	272,4	85,9	127,3
9	281,5	88,3	123,1
10	280,5	83,4	123,3
average	294,26	83,76	122,05
min	272,4	80,7	119,4
max	308,2	88,3	127,3
Spread	35,8	7,6	7,9
Standard deviation	12,19	2,19	2,20

As the third representative, the VSS Vintorez special sniper rifle in 9x39 mm caliber was tested (Figure 8) with the results in Table 4.

Figure 8. VSS Vintorez rifle

Table 4. Measured performance for the VSS Vintorez rifle, caliber 9x39 mm.

Group (10 shots)	Muzzle velocity (m/s)	*LAeq* (dB)	*LCpeak* (dB)
1	283,26	89,20	132,20
2	282,5	90,60	133,30
3	280,26	90,40	132,20
4	283,74	90,40	130,90
5	280,68	90,80	132,70
6	281,91	89,80	131,70
7	283,85	89,50	131,30
8	282,21	90,70	130,50
9	283,44	89,30	131,70
10	282,78	89,20	130,70
average	282,46	89,99	131,72
Min	280,26	90,80	133,30
max	283,85	89,20	130,50
Spread	3,59	1,60	2,80
Standard deviation	1,23	0,66	0,90

For this research, in the fourth case, the AM 20 modular assault rifle platform was used, for which a 9x39 mm caliber barrel was made (Figure 9). Given that the rear cross-section of the case is identical for the 7.62x39 mm, 6.5x39 mm and 9x39 mm calibers, no additional changes to the locking and firing mechanism are required. The idea of the research itself was to examine the possibility of improving the existing AM 20 platform with minimal changes.

The recoil spring was replaced, which is less powerful, but is long enough to safely return and lock the bolt. The powder chamber was modified, which was reduced to fully utilize the pressure of the borrowed powder gunpowder gases, and the percussion spring was replaced. These changes did not affect the functions of the other mechanisms of the rifle.

The existing gunpowder is unknown, so the energy characteristics and dimensions were examined. It was determined that the energy potential is about 3900 J/g and that the shape is a seven-channel cylinder, with a thickness of about 0.176 mm.

Based on the experimental results obtained, it can be concluded that the gunpowder is probably double base. To improve the function of the test rifle, thicker gunpowder with similar energy characteristics was used. Optimal results were achieved using the double-base plate gunpowder NGB-09.

For smooth functional operation, in addition to structural changes to the rifle, the propellant was also replaced, which resulted in higher muzzle velocities, but the maximum pressure is lower than with the original ammunition. The test results are shown in Table 5.

Figure 9. Test Rifle 9×39mm based at the AM 20 modular assault rifle platform

The weapon's silencer is designed by modifying the silencer for the M76 sniper rifle by changing the chambers. In terms of dimensions and volume, it corresponds to the silencer for the VSS Vintorez.

Silencer type 1 is standard silencer for sniper rifle M76 adapted to test rifle caliber. Silencer type 2 is new design of silencer consisting of standard casing (cylinder of M76 silencer) and inner parts that compose four chambers.

Table 5. Measured performance for the Test rifle, caliber 9x39 mm, silencer type1.

Group (10 shots)	Muzzle velocity (m/s)	*LAeq* (dB)	*LCpeak* (dB)
1	309,90	93,60	136,40
2	310,10	94,50	135,00
3	306,80	93,70	130,80
4	312,20	93,40	137,10
5	310,40	93,70	135,70
6	306,70	94,20	136,90
7	308,60	93,40	137,20
8	308,40	95,30	137,30
9	310,00	94,80	136,60
10	307,60	94,00	137,10
average	309,07	94,06	136,01
min	312,20	95,30	137,30
max	306,70	93,40	130,80
Spread	5,50	1,90	6,50
Standard deviation	1,76	0,64	1,97

Table 6. Measured performance for the Test rifle, caliber 9x39 mm, silencer type 2.

Group (10 shots)	Muzzle velocity (m/s)	*LAeq* (dB)	*LCpeak* (dB)
1	305,60	90,20	134,20
2	303,10	91,60	135,30
3	306,50	91,40	134,20
4	306,20	90,40	132,90
5	307,80	90,80	133,70
6	306,30	89,80	132,70
7	303,60	92,50	134,30
8	306,70	91,70	132,50
9	307,00	91,30	134,70
10	303,50	91,20	133,70
average	305,63	91,09	133,82
min	307,80	92,50	135,30
max	303,10	89,80	132,50
Spread	4,70	2,70	2,80
Standard deviation	1,64	0,80	0,91

3. MODEL OF PROJECTILE FLIGHT

Based on the measured results of muzzle velocities, projectile velocities were calculated for all shot cases at distances of 50 and 100 meters. The calculation of the velocities of the projectile was realized according to the modified material point flight model. The governing equations of the model are,

$$m\frac{d\vec{V}_k}{dt} = m\vec{g} + \vec{D} + \vec{L} \tag{1}$$

and,

$$\frac{d\vec{K}}{dt} = I_x\frac{dp}{dt}\vec{x}_0 + I_x p\dot{\vec{x}}_0 + I_y(\vec{x}_0 \times \ddot{\vec{x}}_0) \tag{2}$$

where are aerodynamic forces are,

$$D = \frac{\rho V^2}{2} S C_D(Ma) \tag{3}$$

$$L = \frac{\rho V^2}{2} S C_L(\sigma, Ma) \tag{4}$$

Velocity is defined as,

$$\vec{V}_K = \vec{V}_W + \vec{V} \tag{5}$$

$$V = \sqrt{(u_k - u_w)^2 + {v_k}^2 + (w_k - w_w)^2} \tag{6}$$

where are,

u_k, v_k, w_k components of velocities and

u_w, w_w horizontal components of wind velocities,

$\vec{g}$ gravitational acceleration,

S projectile cross-section surface and,

C_D, C_L aerodynamic coefficients of drag and lift forces.

Angular momentum is defined as,

$$\vec{K} = I_x p\vec{x}_0 + I_y(\vec{x}_0 \times \dot{\vec{x}}_0) \tag{7}$$

where are

I_x, I_y are projectile's moment of inertia,

$\vec{x}_0$ is axial axis of projectile,

$\dot{\vec{x}}_0$ is axial spin of projectile and,

p spin rate of projectile.

Calculated values of the velocities are initial data to determine the terminal parameters (energy and energy density) of the different projectiles at distances of 50 and 100 meters.

4. RESULTS ANALYSIS

Based on the mass of the projectile and the cross-section, the values of kinetic energy and energy density during the interaction with the obstacle were calculated. The values of the quantities mentioned (velocity and energy) are shown in table 7.

Table 7. Performance of the projectiles during interaction

	Velocity [m/s]		Projectile energy [J]		Energy density [J/mm^2]	
	distance [m]					
	50	100	50	100	50	100
AP M84	290	276	193	175	4,03	3,65
HK MP5 SD	281	270	316	292	4,77	4,40
VSS Vinto-rez	273	265	578	544	8,70	8,19
Test rifle	309	299	740	673	11,14	10,43

Projectile energy and projectile's energy density during interaction with an obstacle at 50m and 100m are determined, in relation to velocities, mass and caliber. These parameters are key performances of the presented small arms weapon system.

Figure 10 shows the comparative values of projectile energy at 50m and 100m for the analyzed weapons.

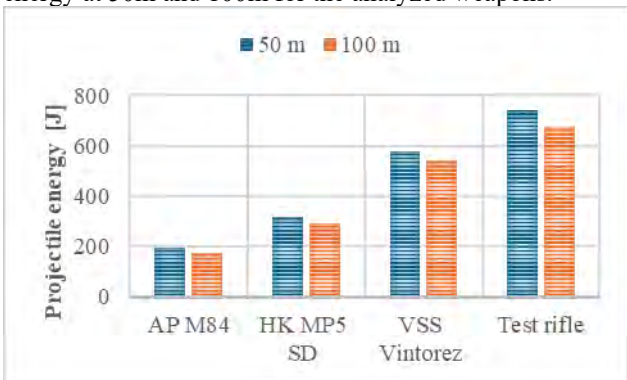


Figure 10. Comparative values of projectile energy at 50 and 100 m

Figure 11 shows the comparative values of the projectile's energy density during interaction with an obstacle at 50m and 100m.

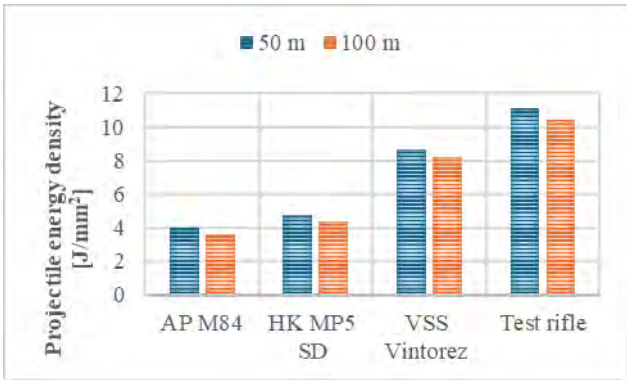


Figure 11. Comparative values of projectile energy density at 50 and 100 m.

Figure 12 shows the comparative values of sound pressure levels for different weapon systems.

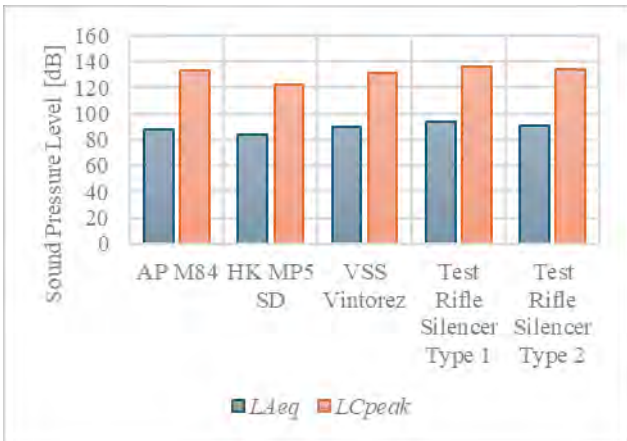


Figure 12. Comparative values of SPL.

The analysis is performed due to verification of the performances of the small arms weapon systems for tactical application in urban conditions and it's evaluating the quality of performance.

The most common representative of this weapon system in most of the world's armed forces is the SMG 9 mm HK MP5 SD, which was selected as a reference system for comparing characteristics.

The characteristics of different weapon models were compared in relation to the reference representative, namely: (a) sound pressure level (SPL) registered with two scales $LAeq$ and $LCpeak$, (b) muzzle velocity level and (c) energy of projectile interaction with an obstacle at the two stated distances.

Automatic pistol 7.65x17 mm M84 showed to have higher SPL on both scales (higher $LAeq$ by 4.82% and higher $LCpeak$ by 9.14%) with higher muzzle velocity by 5.5%. Although the muzzle velocity is higher, due to the significantly lower mass of the bullet (4.6 g vs. 8 g), the energy of the projectile at referential distances is about 40% lower.

The VSS 9x39mm Vintorez rifle was shown to have higher SPL on both scales (higher $LAeq$ by 7.44% and higher $LCpeak$ by 7.92%) with lower muzzle velocity by 4%. Although the muzzle velocity is lower, due to the significantly greater mass of the bullet (15.5 g vs. 8 g), the energy of the projectile at referential distances is about 85% higher.

The Test rifle 9x39 mm with Silencer Type 1 showed that having SPL on both scales (higher $LAeq$ by 12.3% and higher $LCpeak$ by 11.4%) with higher muzzle velocity by 5%. Due to the significantly greater mass of the bullet, the energy of the projectile at referential distances is about 132% higher.

The Test rifle 9x39 mm with Silencer Type 2 showed that it has higher SPL on both scales (higher $LAeq$ by 8.75% and higher $LCpeak$ by 9.64%) with higher muzzle velocity by 4%. Due to the significantly greater mass of the bullet, the energy of the projectile at referential distances is about 131% higher.

To evaluate the new weapon system Test rifles with two types of silencers, a comparison was made in relation to the special sniper rifle VSS of the same ammunition system. The results of the comparison in relation to the VSS rifle are given below.

The Test rifle 9x39 mm with Silencer Type 1 showed that it has higher SPL on both scales (higher $LAeq$ by 4.5% and higher $LCpeak$ by 3.3%) with higher muzzle velocity by 9.4%. The energy of the projectile at referential distances is about 26% higher.

The Test rifle 9x39 mm with Silencer Type 2 showed that it has higher SPL on both scales (higher $LAeq$ by 1.2% and higher $LCpeak$ by 1.6%) with higher muzzle velocity by 8.2%. The energy of the projectile at referential distances is about 25% higher.

To evaluate the performance of different types of silencers, the characteristics of the Test rifle with both silencers were compared. The comparison was made with the newly constructed silencer system Type 2 in relation to Type 1, which represents the existing modified silencer model.

The Test rifle with the new Type 2 silencer showed lower SPL values on both scales. Lower $LAeq$ values by 3.2% and lower $LCpeak$ values by 1.6%. The level of the muzzle velocity was lower by about 1%, as well as the energy of the projectile at referential distances by about 0.6%.

CONCLUSION

The paper presents an analysis of the performance of subsonic weapons for use in urban conditions. Comparisons were made of four types of weapons equipped with silencers. The aim of the research was to present the performance of the Test rifle with two types of silencers and to compare the performance in relation to existing systems. As a reference system, the SMG 9 mm HK MP5 SD was taken, which is the most common and is considered the standard for use in urban conditions.

Based on the presented results, it can be concluded that the weapon systems with the 9x39 mm ammunition system have significantly better performance compared to other representatives. The comparative analysis showed

significant improvements in the performance of the Test rifle in relation to the VSS Vintorez in relation to the muzzle velocity and the energy during the interaction.

If we compare the performance of the two silencer configurations for the Test rifle, it can be concluded that the concept with four chambers showed better results.

Selection of the optimal configuration demands additional tests and analysis of several possible structural solutions of the shot silencer.

Acknowledgment

The paper is a part of the research done within the project of the University of Defence in Belgrade, No. VA-TT/1/24-26.

References

[1] Jerković, D., Hristov, N., Epler, I., Nikolić, B., Aničić, A., Končar, M: *Experimental Analysis of Characteristics of Small Arms Weapon with 9×39 mm Armor-Piercing Subsonic Ammunition System*, Scientific Technical Review, 2025, Vol.75, No.2, pp.3-9.

[2] Rasmussen, P., Flamme, G., Stewart, M., Meinke, D., Lankford, J.: *Experimental Measuring recreational firearm noise*, NOISE-CON 2010 April 19-21, Baltimore, Maryland.

[3] Selech, J., Kilikevicius, A., Kilikeviciene, K. , Borodinas, S., Matijošius, J., Vainorius, D., Marcinkiewicz, J., Staszak, Z.: *Force and Sound Pressure Sensors Used for Modeling the Impact of the Firearm with a Suppressor*, Appl. Sci. 2020, 10, 961; doi:10.3390/app10030961

[4] Rasmussen, P., Flamme, G., Stewart, M., Meinke, D., Lankford, J.: *Measuring Recreational Firearm, SOUND & VIBRATION/AUGUST 2009*; 5 pp

[5] Jokel,C., Yankaskas, K., Robinette, M.: *Noise of military weapons, ground vehicles, planes and ships,* Journal of Acoustics Society in America, 146 (5), 3832–3838, November 2019, https://doi.org/10.1121/1.5134069,

[6] Gurdamar, O., Ozbektas, S., Sungur, B.: *Experimental Investigation of the Projectile type on Sound Pressure Levels Fired with 9 mm Gun,* Khazar Journal of Science and Technology, Volume 6 №1 2022, 96-104, https://doi.org/10.5782/2520-6133.2022.6.1.96 .

[7] Nakashima, A., Hu, Y.: Impulse N oise Exposur e D uring Personal W eapons Testing on an Outdoor Shooting Range, Canadian Acoustics, Vol. 35 No. 3 (2007),192-193.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PRACTICAL APPROACHES TO QUALITY TESTING OF PIONEER MACHINES

DRAGAN STOJADINOVIĆ
Technical Test Center, Belgrade, e-mail: dragan.stojadinovic@vs.rs

ALEKSANDAR ĐURIĆ
Military Academy, Belgrade, e-mail: aleksandar.djuric@va.mod.gov.rs

SNEŽANA JOVANOVIĆ
Technical Test Center, Belgrade, e-mail: sneza.jovanovic.toc@gmail.com

MIROSLAV JOVANOVIĆ
ZF Serbia, Pancevo, e-mail: jovanovic.miroslav.77@gmail.com

SRETEN PERIĆ
Military Academy, Belgrade, e-mail: sreten.peric@va.mod.gov.rs

Abstract: *Pioneer machines, such as diggers, are particularly challenging to test, especially under real operating conditions. This paper presents a practical approach to quality testing of pioneer machines, using a digger as a representative case study, based on procedures applied at the Technical Test Center.*
The study in this manuscript focuses on the implementation of testing procedures required to verify compliance with specified technical characteristics and manufacturer requirements for a digger as a representative machine for the case study. Emphasis is placed on practical testing methods used at the Technical Test Center, including measurement techniques and testing conditions that ensure reliable and repeatable results in real operating environments. The results obtained were based on precise measurements within different testing frameworks and were evaluated against relevant standards and regulatory requirements.
Following the completion of all testing procedures, an overall assessment of the level of compliance with the required characteristics was conducted. In addition, the usability and future potential of the machine were evaluated from the perspective of its design and the analyzed performance characteristics.

Key words: *pioneer machine, testing, digger, exploitation.*

1. INTRODUCTION

The highest hierarchical principle of equipping and developing the Defence system is “creating maximum combat readiness”. That principle is realized with quality means at the level of competitiveness on the world market of armament and military equipment [1]. The degree of quality of an asset intended for the Defence system is evaluated by testing at the Technical Test Center (TTC).

The Technical Test Center is a scientific research institution established by the Serbian Ministry of Defence for testing different parts of armament and other equipment that is going to be part of the official armament of the Serbian Army. The Technical Test Center is testing various parts of regular classic armaments (classical infantry and artillery, missile, aviation, naval, etc.), ballistic protection equipment, different types of electronic equipment and systems, vehicles (combat, non-combat, special), pioneer machines, aircraft, etc. Tests are performed in laboratory, testing ground, and field conditions. All the measurements and tests performed and being conducted at the TTC are defined by both international and domestic standards. Likewise, the TTC is making some new standards and new test methods based on experience and modern achievements in measurement techniques and new measurement equipment. The measured values are being analyzed and compared both with the required characteristics specified in the standard and with the declared characteristics of the manufacturer, after which a decision is made on official use in the Serbian Army [2].

The pioneer machines are machines made for hard work in the construction industry, road construction, snow cleaning, and all other work that needs a lot of energy and force to be used in it [3]. One of the pioneer machines that is commonly used in the Serbian Army units for various heavy works is the digger. It is a self-propelled combined excavator-loader, which is being used for heavy land working and loading of material, material excavation, and heavy material moving with an integrated loader bucket as a working tool [4]. The machine has a lot of different additional equipment that gives us different possibilities: to break materials, dig holes in different ground types with a diameter depending on the used drilling tool, strip different layers of land, asphalt, and concrete, and meet

Corresponding author: D. Stojadinović
E-mail: dragan.stojadinovic@vs.rs

DOI: 10.5937/OTEH26021S

various other demands as it is stated in [4] and [5].

Within the framework of pioneer machine testing at the TTC, a comprehensive evaluation is conducted with the aim of determining the level of compliance with all required technical characteristics and tactical/technical requirements defined by the customer (in this case, the Serbian Army). The testing includes an assessment of the machine`s usability under real exploitation conditions, taking into account specific operational requirements imposed by the user, as well as an evaluation of the modernity and future viability of the applied technical solutions, both at the level of individual subsystems and the system as a whole [4]. The majority of the tests are carried out with a focus on determining the exploitation and maneuvering capabilities of the machine.

The object of examination in this manuscript is a digger which is now in use in the Serbian Army, model JCB 4CX (Figure 1), with the following tools:
- loading bucket,
- digging spoon,
- soil drill,
- fork,
- hydraulic breaker.

Figure 1. Test object, digger JCB 4CX

2. PIONEER MACHINE TESTING

For this tested pioneer machine, first the Test Program is prepared in order to determine all parameters and characteristics that will be checked. The basis for the Testing Program is the declared characteristics of the manufacturer, the tactical/technical requirements of the institution that orders the testing (the Serbian Army in our case), as well as new standards and methods for testing. All tests will be performed according to the adopted Test Program.

The purpose of the test is to assess:
- Satisfaction of the tactical-technical requirements of users,
- Compliance of the declared characteristics with the tactical/technical requirements,
- Usability of the asset in real conditions,
- Modernity and perspective of the applied solutions,
- Ease of handling and maintenance.

Generally, mechanical testing of excavators consists of four stages, as shown in the block diagram in Figure 2.

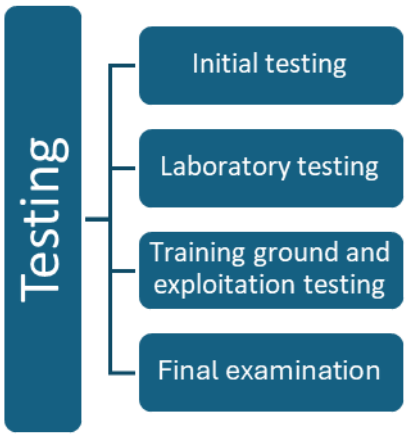


Figure 2. Test phases

2.1. Initial testing

Initial testing is carried out with the aim of determining whether there are basic prerequisites for conducting the test. First, the completeness of the construction documentation and identification data of the asset and basic subassemblies are determined, as well as a comparison of basic parameters with available documentation. The validity of attestations and certificates is checked. The general condition and proper functioning of the device are visually determined.

The intention of this part of testing the pioneer machine is as follows:
- recognition and identification of all constituting parts of the digger, machine sets and assemblies according to the technical documentation of the machine provided by the manufacturer,
- checking of the engine and engine systems, including the parts they are made of, specifically measurement of some engine and engine system parameters (air suction system, engine fuel supply system, system for engine lubrication, exhaust system, engine cooling system, checking the level of all working fluids in the engine, heating and ventilating system of a digger, etc.),
- transducer installation for some engine parameters (temperature of oil, cooling fluid etc.),
- checking the built-in parts of the digger transmission, tyres, measuring tyre pressure, all fluids and fluid levels in transmission sets, and checking the functionality of the digger transmission, etc,
- identification of the parts of the digger's braking, steering, and hydraulic systems (hoses, pumps and motors, small additional tools attachment set, etc.), fluid levels in all mentioned systems, etc.;
- Electromagnetic compatibility (EMC) characteristics (in documentation);
- basic checking of the functionality of all digger systems and simulation of work with machine tools.

Following the completion of all initial tests, it is necessary to conduct, at the end of the initial tests, a trial run of approximately 20 km on an asphalt road, in order to evaluate the performance and functional behavior of all components, assemblies, and the machine under

operational conditions [4].

2.2. Laboratory testing

Laboratory testing is customized for each individual machine and performed with respect to the specific characteristics of the test object. It generally consists of several stages and associated measurements, including:

- Geometric parameters of the machine [6],
- Digger mass characteristics – total mass of the machine, machine weight distribution by wheels and axles, machine center of gravity [7],
- Backward movement,
- Ergonomic noise characteristics – internal noise level in the cabin, external machine noise level, siren noise level [8],
- Ergonomic temperature characteristics in extreme environmental conditions – at temperature levels from -30ºC to +50ºC, checking the engine start at low (-12ºC) and high (+46ºC) ambient temperatures [9], [10],
- Determining Predicted Mean Vote (PMV – seven-point thermal scale that shows to what extent the comfort temperature is achieved in the real environment, namely: -3 (icy), -2 (cold), -1 (moderately cold), 0 (comfortable), 1 (moderately warm), 2 (warm), 3 (hot)) and Predicted Percentage of Dissatisfied (PPD - predicts the percentage of people dissatisfied with the thermal environment, that is, the percentage of people who will feel thermal discomfort in a certain real environment) index at low external temperature (-16ºC) and normal external temperature [11], [12],
- Characteristics of all electrical wiring systems and electrical equipment, voltage levels, current intensity, short circuit protection, insulation, alternator voltage level [13], [14],
- EMC characteristics [15],
- Corrosion protection [16],
- Safety features,
- Features related to environmental protection.

2.3. Training ground and exploitation testing

Training ground and exploitation tests include the examination of functional and other characteristics in real conditions of exploitation, which derive from the name of the means being tested, as well as the set requirements. The main goal of this test phase is to get as close as possible to the realistic working conditions of the machine. Polygon and exploitation tests involve some of the most intensive testing conditions we can put on a single machine. These tests include:

- Dynamic characteristics testing [17], [18], [19],
- Intensive driving testing of the machine in a few specific road conditions [20],
- Reliability, efficiency and durability of work tools,
- Accuracy, precision and life expectancy, etc.,
- Safety and resilience.

Testing the mechanical durability by driving on different types of roads is a mandatory part of this test phase, and when it comes to machines, it is also mandatory to test the work with all the tools that are mounted on it, also in different conditions.

The most important dynamic characteristics of the machine are:

- Stability and passability (overcoming longitudinal and lateral slope, clearance) [21],
- Handling performance (turning radius, steering sensitivity),
- Braking performance (stopping distance),
- Startability (acceleration and maximum speed).

Within the mentioned object of testing, dynamic characteristics according to [17] are being determined:

- Evaluation of the machine's capability to overcome longitudinal and lateral inclinations (gradients), including stability and mobility performance under such conditions,
- Wheel radius (static, dynamic);
- Evaluation of the machine's handling characteristics, including minimum turning radius and steering response sensitivity under defined operating conditions [22],
- Testing the braking system, braking distance measurement in driving testing [23],
- Highest and lowest velocity in manual and automatic mode in all gears,
- Acceleration in manual and automatic mode in all gears and acceleration from minimum to maximum velocity,
- Vehicle resistance using the coast method, etc.

Intensive driving testing of the machine in a few specific road conditions is a very important part. This is specific for a vehicle and/or machine by itself. For this object of testing - a digger, it includes driving of the machine for around 500 km, under a few different road conditions: about 250 km on plain and hill ground road and about 250 km under heavy off-road conditions.

While driving, some basic indicators are being monitored, where the most important are:

- Test road conditions, ground and meteorological conditions while driving for every day of testing,
- Total distance driven in kilometers per day and in total,
- Driving time, per day/total, time of standing (non-work time),
- Failures, fractures, and possible crushed parts of the machine construction that occur while testing is being done, with the number of kilometers at which the failure occurs being recorded,
- Regular maintenance, adjustment, and servicing while testing, with the number of kilometers being recorded,
- Durability of parts and sets in transport.

Reliability and efficiency of digger working tools are a special phase of testing. In this phase, work with the digging spoon is performed for a total of 25 engine working hours, while excavating, laying-off, and loading land, and working with the loading spoon is performed for a total of 25 engine working hours [4].

The objective of this testing is to verify and evaluate the following parameters:
- maximum excavation depth,
- ground reach of the excavator arm,
- kinematic and functional characteristics of the bucket, including digging speed, bucket width, and load-carrying capacity, as well as the operating speed of the loading shovel,
- functionality of the hydraulic control system, maximum working pressure within the hydraulic installation, condition of hydraulic hoses, and integrity of the system with respect to leakage and overall operational reliability.

Parallel with testing, the quality of oils is being checked, together with digger fuel consumption (in all different road and working conditions) and engine characteristics when working under maximum load, high external temperature levels, etc.

If auxiliary attachments are integrated into the machine (e.g., hydraulic hammer, concrete saw, drilling, etc.), additional testing procedures are required, including:
- verification of functional performance over a continuous operating period of 5 hours during the processing and fragmentation of various materials,
- assessment of drilling capability for boreholes with diameters up to 950 mm,
- evaluation of performance in the removal of asphalt and concrete layers with widths up to 400 mm and depths up to 120 mm,
- testing of cutting efficiency and the capability for demolition of concrete structures,
- evaluation of performance in auxiliary operations such as snow removal, etc.

2.4. Final examination

Final tests are carried out with the aim of determining whether there are any major failures or major deviations from the initial machine characteristics. Usually, all the parts of initial testing are performed again, but this time with the goal of determining major differences between the start and the end of the machine testing and deviations from the machine construction documentation.

3. ANALISE OF TEST RESULTS

3.1. Results of initial testing

At initial testing of a digger the conclusions were:
- All parts and sets are made according to the documentation of the machine provided by the machine manufacturer;
- All parts and sets are working correctly,
- All the fluids in assemblies are at proper levels that are set by the manufacturer. Also, the working fluids in the machine are filled according to technical documentation and maintenance manual, provided by manufacturer of the machine,
- Digger start-up is smooth and continuous, without abrupt or irregular dynamic responses,
- Transmission is working properly in manual and automatic mode and when moving forward and in reverse, in every gear, etc;
- The braking system is working, as well as the steering system,
- All digger sets are working without irregular responses,
- Additional tools are working properly,
- Digger is equipped with appropriate safety systems, including rollover and falling object protection, an adjustable steering mechanism, adequate ventilation, a safety belt, an audible warning device, and other relevant operator safety features, etc.

3.2. Conclusions of laboratory testing

The measured dimensions (including transport height, width, length, bucket width, etc.) comply with the required specifications, with negligible deviations of less than 5%, which are considered acceptable within the measurement tolerance limits, as stated in [4].

According to [4], the angles of the rear wheel turning are from 15° to 18°, which is acceptable. Also, the total mass and axle mass distribution are according to the documentation provided by the manufacturer.

The results of the digger cabin noise level measurements are given in Table 1.

Table 1. Digger cabin noise level

Noise level	Digger cabin	Directive 2003/10 [24]
Close windows	79,5 dB	Above 72 dB
Open windows	84 dB	Below 86 dB

The low and high temperature testing is performed according to [9] and [10], with all the required test preparations and conditions. The engine has successfully started both at high external temperature (+46ºC) and low external temperature (-12ºC) in the cooling and heating test chambers.

Together with high and low external temperature digger testing in cooling and heating chambers, the efficacy of the cabin heating system has been tested [11], [12], as shown in Table 2.

Table 2. Digger cabin heating system efficiency

Digger cabin	PMV	PPD
Measured	-0,5	10%
Criteria	$-0{,}5 \leq PMV \geq +0{,}5$	$\leq 10\%$

The electrical installation of the machine was characterized as follows:
- the conductors are adequately protected against mechanical influence,
- the alternator provides stable and reliable power supply, ensuring proper battery charging and maintenance,

- the minimum measured insulation resistance was 400 MΩ. Narrow-band interference at high frequencies slightly exceeded 400 MHz, with a maximum permissible deviation of 5 dB,
- prior to operation, the measured voltage drop ranged from ΔU = 0.30 V to ΔU = 1.25 V. After a 500 km test drive and 50 engine operating hours under working conditions (earth excavation), the voltage drop increased to a range of ΔU = 0.70 V to ΔU = 2.25 V, indicating a significant increase compared to the initial measurements. The system maintained a constant voltage of 14.4 V across an engine speed range from 500 rpm to 3000 rpm, with regulation provided by the alternator voltage controller.

3.3. Conclusions of training ground and exploitation testing

Dynamic characteristics of the machine were:

- Equal digger velocity at all gears, both in forward or reverse, except 4th gear, 38.9 km/h in reverse/ 40 km/h in forward motion (digger meets the demand for a maximum velocity of 40 km/h/or in the range of 40 km/h to 50 km/h),
- Satisfactory tire rolling radius according to the technical documentation provided by the manufacturer;
- The braking performance of the machine was assessed as satisfactory. Type "0" brakes with engage/disengage clutch demonstrated adequate efficiency, as the measured maximum stopping distance was below the permissible limit. The performance of Type "I" service brakes was also evaluated as satisfactory. The parking (hand) brake was capable of holding the machine stationary on the maximum tested gradient of 30%.

Testing of the machine in intensive driving conditions was performed similarly to [4] on plain-ground roads and under heavy off-road conditions. The average velocity of 14 km/h on plain roads and an average velocity of 15,5 km/h under heavy off-road conditions was rated as positive for this type of machine.

Functional performance and operational safety under working conditions were evaluated during excavation operations using both the backhoe bucket and the loader bucket over 50.6 engine operating hours, also similarly to [4]. The backhoe bucket was utilized for earthworks in soil of category II, while the loader bucket was also operated under category II soil conditions. During the test, the temperature of the engine fluids was monitored. The results obtained were within acceptable limits and were assessed as satisfactory.

Some characteristics of the digger, measured within exploitation tests, are given in Table 3:

Table 3. Some digger characteristics

Digger	Measured	Demand
Max. digging depth with spoon	5970 mm, fulfill	More than 5800 mm
Digging spoon velocity	0,24 m^3	-
Lifting Force of digging spoon, extracted telescope	1200 kg	
Lifting Force of digging spoon, sucked telescope	1550 kg	
Loading spoon velocity	1,22 m^3	
Lifting Force of forks integrated with spoon	1000 kg	
Total rate of flow	152 l/min	
Max. working presure	226 bar	

4. CONCLUSION

The combination of a backhoe and loader on the same engineering machine is the most optimal solution for the needs of the army. This model has a multi-purpose loading hydraulic opening bucket, according to the 6-in-1 system, with bucket blade protection and an additional fork. The machine has a power shift transmission and a high level of comfort. The construction of the digger ensures quick assembly of all additional tools (hydraulic hammer, spiral drill, fork, hydraulic breaker, concrete grinder, etc.), which is of particular importance for the purpose of the army pioneer units.

However, in terms of construction and workmanship, minor defects were observed. The measured turning angle of the wheels is less than required. The noise in the cabin is higher than allowed and a voltage drop was also found in some parts of the electrical installation, which directly affects the driving and working conditions in operational conditions.

Based on the comparison of the obtained results with other solutions of this type of pioneer machines, as well as based on the previous experience of researchers at the Technical Test Center, the summary view is that the digger meets the set requirements and is a promising product in terms of intended use.

This paper presents one approach of the TTC when testing a pioneer machine. With the results obtained by precise measurements, the customer can come to a conclusion as to whether a certain product meets the conditions for use and procurement or not. In this specific case - testing of the digger, the results are mostly positive, with minor defects, which do not significantly affect its operational use. We believe that this approach is expedient and recommended when purchasing this type of machine. Certainly, there is always the possibility of additional tests and checks, all in accordance with the needs of the client.

References

[1] BRKLJAC, N.: *Quality in the function of combat readiness maximization*, Vojnotehnički glasnik/Military Technical Courier, 54(3), pp. 342-347 (in Serbian). Available at: https://doi.org/10.5937/vojtehg0603342B, 2006.

[2] http://www.toc.vs.rs/ - TOC website, accessed April 10, 2026.

[3] MARKOVIĆ, V.: Construction Machines for earthworks (in Serbian), Scientific Book, Belgrade, COBISS-ID: 6028293, 1975.

[4] JOVANOVIĆ, S., ĐURIĆ, A., REPAJIĆ, D.,: *One Approach in Testing Pioneer Machines*, Heavy Machinery Conference, Kraljevo, ISBN 978-86-82631-58-3, 2011.

[5] LAZAREVIĆ, D.: *Engineering Machines 1* (in Serbian), Military Academy, Belgrade, COBISS-ID: 114191116, 2003.

[6] SORS 9284, *Determination of Geometric Parameters of Patency*, 1998.

[7] SORS 0027, *Mass Characteristics of Military Vehicles*, 2005.

[8] ISO 6394, ISO 6396, *Earth-moving Machinery – Determination of Emission Sound Pressure level at Operator's Position*, 2008.

[9] SORS 0142, *Test Method. Starting Characteristics of Internal Combustion Engines at Low Ambient Temperatures*, 2003.

[10] SORS 1026, *Characteristics of Engines and Engine Devices. Examination of the Starting Characteristics of the Engine at High Ambient Temperatures*, 1986.

[11] SORS 8419, *Examining the Thermal Comfort of Working Areas of Military Vehicles*, 2002.

[12] ISO 7730, *Modern Thermal Environments - Determination of PMV and PPD Indices and Definition of Thermal Comfort Conditions*, 2008.

[13] SORS 5574, *Electrical Installation and Equipment. Methods of Testing Electrical Resistance*, 2003.

[14] SORS 6711, *Combat and Non-combat Vehicles and Tools. Electrical Installation. Methods of Testing Voltage Drops*, 1986.

[15] ISO 13766, *Earth-moving Machinery – Electromagnetic Compatibility*, 2006.

[16] SORS 2348, *Black Oxide Coating on Iron and Steel*, 1999.

[17] SORS 0028, *Examination of Speed Characteristics of Wheeled Vehicles*, 2002.

[18] SORS 9156, *Traversability Characteristics of Military Vehicles,* 1998.

[19] SORS 9277, *Examination of Ability to Overcome Artificial Obstacles*, 1998.

[20] SORS 9422, *Exploitation of Driving Tests*, 2018.

[21] SORS 9156, *Traversability Characteristics of Military Vehicles. Longitudinal and Transverse Slope*, 1998.

[22] SRPS EN 12643, *Earth-moving Machinery - Tire-driven Machinery - Control Requirements*, 2009.

[23] SORS 9303, *Characteristics, Dynamic, Military Vehicles. General Technical Regulation for Final Tests of Brake Systems*, 1998.

[24] Directive 2003/10/*EC of the European Parliament and of the Council / Rulebook on Preventive Measures for Safe and Healthy Work When exposed to Noise*, SG RS 96/11, 78/15 and 93/19.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

IMPLEMENTATION AND EXPERIMENTAL VERIFICATION OF ENVIRONMENTAL SAFETY SENSOR DEVICE FOR FUZES IN AERIAL BOMBS

SLOBODAN PELEMIŠ
Singidunum University, Vlatacom Institute, Belgrade, slobodan.pelemis@vlatacom.com

NIKOLA LATINOVIĆ
Vlatacom Institute, Belgrade, nikola.latinovic@vlatacom.com

MLADEN VEINOVIĆ
Singidunum University, Belgrade, mveinovic@singidunum.ac.rs

Abstract: *In this paper, a methodology for the implementation of an environmental safety sensor device as part of a fuzing system for aerial bombs will be presented. Complete hardware and software for this system architecture will be described herein.*
Environmental sensors have a wide range of applications in industry, military technologies, IoT, and many more. These are usually devices that can measure various physical parameters of the environment and convert data into electrical signals suitable for use with electronic devices. In military applications, in weapon systems, the most important factor is safety during the handling and use of explosive products. In this particular case, the application of the environmental sensors in the fuzing systems for aerial bombs is crucial for the safety of pilots and aircrafts.
The test results of the implemented device will be shown as proof of functionality under the given conditions. Software is written in C language as the operating system of a microcontroller used as the main control unit of the entire system.
The core contribution of this work is the analysis of safety features of a device that generates the appropriate signals as conditions for the aerial bomb fuze's performance of its arming process.

Keywords: *safety sensor, environmental conditions, electronics, aerial bomb, C language.*

1. INTRODUCTION

Aerial bombs detonation is controlled by the action of a fuze. A fuze is a device that causes the detonation of an explosive charge at the proper time after certain conditions are met. Part of the fuzing system is also a product called initiator mounted in the middle of the bomb between lugs as shown in Figure 1. [1]

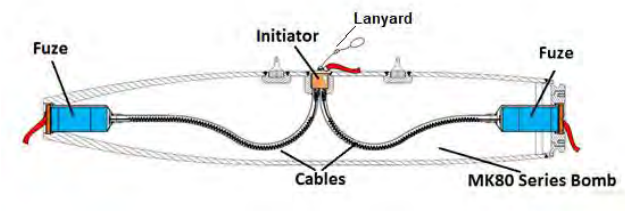


Figure 1. Position of the fuzes and initiators in aerial bomb

The initiator is a cylindrically shaped metal component that is installed in the bomb charging well.
The initiator contains two essential components:

- ***Launch sensor,*** MP1 (switch): which in the active position (closed) connects the battery to the electronic block of the fuze and to the initiator;
- ***Environmental sensor*** with appropriate comparator. This component provides data, designated as MP2, about the pressure of the air stream, indicating that the aircraft has released the bomb and the bomb is flying to the target.

A dedicated hardware-software architecture has been developed for the realization of the proposed sensor device for safety in air-bomb fuzing systems.
The system relies on the integration of a microcontroller unit (MCU) and a differential pressure sensor to ensure that the arming signal is generated only after meeting the given environmental conditions.

Environmental sensor does not serve to activate the bomb on the target, but as a safety device that prevents the fuze from operating until the bomb has safely left the vicinity of the aircraft. [1], [2]

Safety role means that the sensor will generate arming signal when the pressure of air stream is greater than 0.2 psi. This signal is monitored by MCU in the electronic block of the fuze during the bomb flying.

Advantage of this integrated software in a relation with pressure sensor is in the following: when we are comparing system used in AUFK M-91 fuze, with system presented in this paper, we can see that AUFK M-91 fuze

Corresponding author: S. Pelemiš;
E-mail: slobodan.pelemis@vlatacom.com
DOI: 10.5937/OTEH26022P

in order to be armed, must meet the electrical condition, i.e. to receive a certain electric pulse of nominal value of 24V from the plane. It represents a safety requirement and as such would not be acceptable to the modern 1316 safety standard used today. [3]

In this paper, by introducing an initiator based on the differential pressure sensor, we have two safety conditions instead of one. First one represents the opening of the switches when dropping the bomb in form of MP1 signal. Second one is a physical component, i.e. the measurement of differential pressure, which is directly proportional to the speed of the air flow, in real time conditions shown as MP2 signal.

This creates two independent safety conditions that form the basis for the fulfillment of standard 1316.

2. HARDWARE DESIGN AND SIGNAL CONDITIONING

Environmental pressure sensor is located inside the steel housing of the initiator and is activated during the flight of the aerial bomb.

Electronic assembly of complete system which includes mentioned sensor is shown in **Figure 2**.
Characteristics of sensor are given in **Table 1**.

Table 1. Pressure sensor characteristics

Characteristic	Symbol	Min	Typ	Max	Unit
Pressure Range	P_{OP}	0	-	50	kPa
Supply Voltage	V_S	4.75	5.0	5.25	V_{dc}
Supply Current	I_O	-	7.0	10	mAdc
Min. Pressure Offset (0 °C to 85 °C), Vs=5 V	V_{off}	0.088	0.2	0.313	Vdc
Full Scale Output (0 °C to 85 °C), Vs=5 V	V_{FSO}	4.587	4.7	4.813	Vdc
Full Scale Span (0 °C to 85 °C), Vs=5 V	V_{FSS}	-	4.5	-	Vdc
Accuracy (0 °C to 85 °C)	-	-	-	±2.5	$\%V_{FSS}$
Sensitivity	V/P	-	90	-	mV/kPa
Response Time	t_R	-	1.0	-	ms
Output Source Current at Full Scale Output	I_{o+}	-	0.1	-	mAdc
Warm-up Time	-	-	20	-	ms
Offset Stability	-	-	±0.5	-	$\%V_{FSS}$

Upper PCB contains differential pressure sensor and ESD protection diodes.

Lower PCB contains pressure switches which are responsible for providing battery power supply to the fuze electronic block.

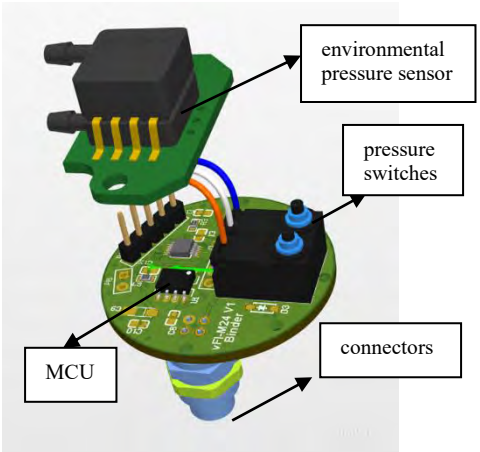


Figure 2. Electronic PCBs of the initiator

The system relies on the integration of a microcontroller unit (MCU) [4] and a differential pressure sensor to ensure that the arming signal is generated only after meeting the given environmental conditions.
Electrical block scheme of complete system is shown in Figure 3**.** [5]

The main purpose of the system is to measure differential pressure, process the sensor signal inside the MCU, and generate a digital output signal when the measured pressure exceeds a predefined Threshold. [6]

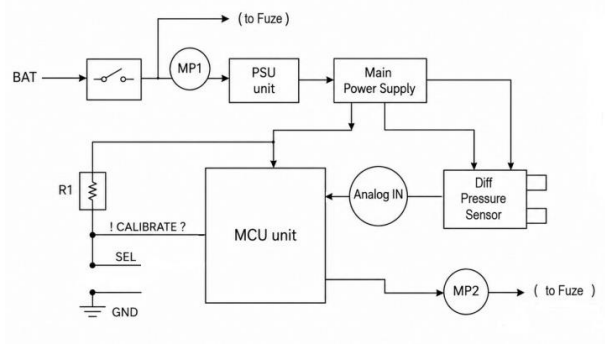


Figure 3. Electrical block scheme of complete system

The system is powered by the battery input (**BAT**). After passing through the main power switch, the supply voltage is available at signal **MP1** which is first signal to be achieved. **PSU Unit** represents preliminary power conditioning stage. Main Power Supply generates the regulated voltages required by the MCU and the differential pressure sensor.

The regulated outputs from the Main Power Supply provide stable operating voltages for the MCU unit and the differential pressure sensor. This arrangement ensures reliable operation regardless of battery voltage variations.

The Differential Pressure Sensor measures the pressure difference between its two pressure ports.

The sensor produces an analog voltage proportional to the measured differential pressure. This analog signal is routed to the MCU through the Analog IN interface.

The MCU continuously samples this signal using its internal Analog-to-Digital Converter (ADC), converts the analog voltage into a digital value, and evaluates the measured pressure against predefined limits.

The calibration circuitry consists of:

- R1 pull-up resistor;
- CALIBRATE input;
- SEL input;
- Ground reference.

These signals allow the system to enter a calibration mode during startup.

When the CALIBRATE line is active, the MCU performs a sensor offset calibration procedure and stores the calculated offset values into non-volatile memory. This compensates for sensor tolerances, temperature effects, and manufacturing variations. [7]

Regarding the output section, it is important to note that the **MCU** generates a digital output signal that is available at MP2, which indicates wheather the measured pressure exceeds the programmed Threshold.

3. SOFTWARE DESIGN AND IMPLEMENTATION

The software support, implemented in the C language, directly controls the states of the hardware components shown in the schematic on **Figure 3**. [8]

Sensor expected measurement range indicated in the experimental results equates to the value of V_{off} + 620.7 mV ± 5% of measurement inaccuracy. The value of 620.7 mV is derived from the formula presented by the device vendor:

$$V_p = V_{off} + \frac{90}{0.145} \cdot P\ [mV] \qquad (1)$$

where V_p is the output voltage in [mV], V_{off} is voltage offset in [mV], and P is pressure (difference) in [psi] which produces the output voltage V_p.

3.1 SOFTWARE CODE WRITTEN IN C LANGUAGE

Following lines show the code written for check of the 0.2 psi pressure value of the sensor. [9], [10]

```
/*
 * Pressure sensor equation: Vp-Voff = (90/0.145)*P [mV]; P is in Psi units
 * for 0.2Psi Vp-Voff = 124mV
 * if using 12bit A/D conversions 1LSB is 1.22mV so threshold is equal to 102
 */

#define MP2_THRESHOLD12b  102

//Functions prototypes
unsigned char MP2_Read2B (void);
void writeOffsetToMem2B (unsigned char tempHigh, unsigned char tempLow);
void MP2_Calibrate2B (unsigned char*memLow, unsigned char *memHigh);

//Main program
void main (void)
{
   unsigned char  memLow=0, memHigh=0;
   int i=0, j=0;
   P04_PUSHPULL_MODE   //P0.4 output push pull  -- MP2
   P01_INPUT_MODE;        //P0.1 input mode -- CALIBRATE
  P06_PUSHPULL_MODE;          //led on initiator board

   P04=0;
   P06=0;
   if(P01)
   {
            while(1)
            {
                     if (MP2_Read2B())
                     {
                              P04=1;
                     }
                     else
                     {
                              P04=0;
                     }
            }
   }
   else
   {
            P06=1;
            MP2_Calibrate2B(&memLow, &memHigh);
            writeOffsetToMem2B(memHigh, memLow);
            P06=0;
            while(1);
   }
}
```

Software support, implemented in C language, directly manages the states of the hardware components shown in the diagram in **Figure 3**.
At the very beginning of the execution of the *main()* function, the input-output pins of the microcontroller is initialized as follows:

- *Port P0.4* is configured in push-pull operation mode. According to the scheme on **Figure 3**, this pin directly controls the MP2 output signal to the fuze.
- *Port P0.1* is set to input mode of operation and corresponds to the line *!CALIBRATE?* from the electrical diagram.
- *Port P0.6* is configured as a push-pull output to control the LED on the initiator board, which serves as a visual status indicator during calibration.

After initialization, the software checks the state of pin *P0.1(if(P01))* to determine the branching of the algorithm:

- **Detection and assurance mode (Normal operation)**: If pin *P0.1* is at a high logic level (calibration switch is open), the system enters an infinite *while(1)* loop where the *MP2_Read2B()* function is continuously called. This function reads the value from the Analog IN position

(pressure sensor output). In the code, the safety threshold *MP2_THRESHOLD12b* is defined with a value of 102. This value directly corresponds to a physical pressure of 0.2 psi (that is, a voltage shift of 124 mV above the offset, taking into account the 12-bit resolution of the ADC where 1 LSB is 1.22mV). If the differential pressure read exceeds the defined threshold, the software sets pin P0.4 to a high logic level *(P04=1)*, which activates the output to the igniter via MP2. Otherwise, the output remains inactive *(P04=0)*, ensuring system security.

- **Calibration mode**: If pin *P0.1* is low at startup *(else branch)*, the LED is activated (*P06=1*). The software then calls the function *MP2_Calibrate2B (&memLow, &memHigh)* to measure the current offset of the pressure sensor at rest, and then writes these values permanently to memory via the function *writeOffsetToMem2B(memHigh, me*mLow). Upon completion of the calibration, the LED turns off (*P06=0*), and the microcontroller enters a waiting state (*while(1)*), blocking further operation of the igniter until the system is restarted in normal mode.

With this synergy of hardware lines and software conditions, a high resistance to unwanted activation of the aerial bomb was achieved, because the arming signal cannot be generated without an adequate change in the environmental pressure that follows the specific flight trajectory of the aircraft.

4. EXPERIMENTAL TESTS WITH RESULTS AND FUNCTIONALITY PROOF

In order to verify developed software which is implemented on MCU, experimental tests are performed. [11]

Equipment necessary for these tests is:

- pneumatic calibration pump with connected digital pressure gauge;
- initiator electronics test box;
- power supply 10.8 V DC;
- voltage meter.

Pneumatic calibration pump with digital pressure gauge shown in **Figure 4** is necessary for simulating air stream pressure which will occur in real-time working conditions during air bomb flight.

Figure 4. Pneumatic calibration pump with digital gauge

By using the appropriate adaptor and a hose, pressure port of the sensor of UUT and the output of pneumatic calibration pump must be connected as shown in Figure 5. UUT in this case is pressure sensor located inside of the initiator housing.

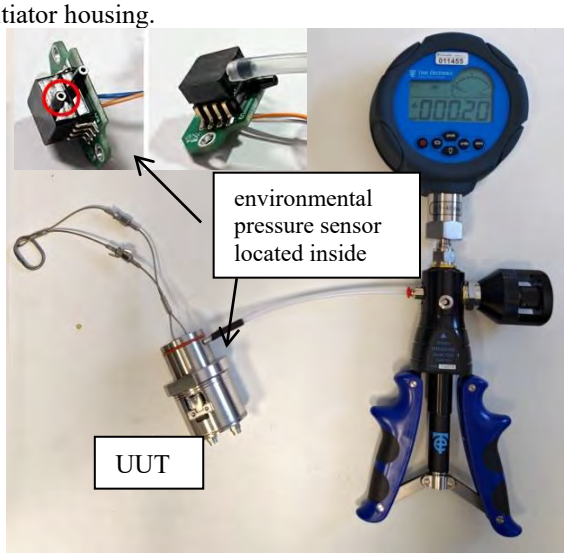


Figure 5. Pump connected with UUT sensor

First part of the test setup is shown on Figure 5. To complete test, we must have also connected power supply, voltage meter and initiator test box. Initiator test box is used to show when MP1 and MP2 signals are activated. Rest of the setup we can see on Figure 6.

Figure 6. MP1 signals activated

In above Figure, it is demonstrated that with air stream pressure lower than 0.2 psi (see display on calibration pump), we have only MP1 signals active (MP1 diodes

glows red, and MP2 diodes don't). This means that when air bomb is released on active mode, switch MP1 is closed, the battery connects to electronic block and arming process started. Environmental sensor still didn't detect air stream and didn't produce any analogue output signal. This is implemented through the C language written code.

Next step is to produce pressure of 0.2 psi or bigger which happens during bomb flight. [11]

We can see pressure value manually defined on pump on **Figure 7**.

Figure 7. MP2 signals activated

When the pressure of the air stream reaches 0.2 psi or above, pressure sensor produce signal MP2=TRUE. This signal is monitored by the microcontroller in the electronic block of the fuze during the bomb moving. The microcontroller starts monitoring this signal when 0.9 of the arming time has elapsed, i.e. at the beginning of the last 10% of the arming time. The sensor will generate an MP2 signal TRUE only if the pressure of air stream is greater than 0.2 psi. (for example if the arming time is preset at 10 seconds, in order for the fuze to arm, this presure must be greater than 0.2psi after 9 and before 10 seconds of arming time have elapsed). **Figure 7** shows that, with the appropriate pressure applied, beside the MP1 diodes, the MP2 diodes on the test box also glow, which indicates the start of the arming process. [11]
In order to prevent influence of noise and interference the microcontroller checks signal MP2 three times after the completion of the 90% of arming time. In case the signal is present during all of these checks, microcontroler concludes that the bomb is moving away and distance from the aircraft is safe enough for the arming process of the fuze to be completed.. Conversely, if the MP2 signal is not TRUE during the last 10% of arming time (3 consecutive readings), arming process will be stopped and arming time resets. The fuze will not be armed.

5. CONCLUSION

This paper has presented the implementation and experimental verification of an environmental safety sensor device designed for aerial bomb fuzing systems. By integrating a microcontroller unit with a differential pressure sensor, the developed hardware-software architecture successfully addresses the strict safety demands critical to military applications.

The embedded software, written in C language, provides reliable signal processing and introduces a dedicated calibration routine that effectively compensates for sensor tolerances and environmental variations.

Through the experimental laboratory setup which includes a pneumatic calibration pump and a dedicated test box, we successfully validated the system's functionality under simulated flight conditions. The testing confirmed that the crucial arming signal (MP2) triggers only when the airflow pressure safely crosses the 0.2 psi threshold.

It is shown that comparing to other similar products used in military, we have two safety conditions ensured (explained through MP1 and MP2 signals) which is advantage regarding the modern military standard 1316. Other fuzes such as for example AUFK type has one condition in form of electrical impulse variable.

Ultimately, the core value of this design lies in its contribution to pilot and aircraft safety. By tying the arming sequence directly to a precise, verifiable change in environmental pressure along the flight trajectory, the system achieves excellent resistance to accidental or premature activation.

References

[1] U.S. DEPARTMENT OF DEFENSE, MIL-STD-1316F: *Safety Criteria for Fuze Design*, Washington, DC, USA – DoD, 2017.

[2] DEURA,N., DOMAZET,D., NIJEMČEVIĆ,S.: *Fuze vAF-M17 Electronic block test software*, in Proc. 10th Scientific Conf. Defensive Technologies (OTEH), Belgrade, Serbia, pp. 121-127.

[3] KRUŠIK,H.,K.,*Upaljači,* https://krusik.rs/images/pdf/UPALJACI.pdf, page 26

[4] PEPLOW,R., SHENOUDA,D.: *Digital System Design – Use of Microcontroller,* River Publishers

[5] OH,J.S., HAN,Y.M., CHOI,S.B., NGUYEN,V.Q., MOON, S.J.: *Design of a one chip board microcontrol unit for active vibration control of a naval ship mounting system*, IOP Publish. doi: 10.1088/0964176/21/8/087001.

[6] ŽIVKOVIĆ,M., DOMAZET,D., LATINOVIĆ,N.: *Fuze vAF – M17 microcontroller software*, in Proc. 10th Int. Scientific Conf. Defensive Technologies (OTEH), Belgrade, Serbia, 2022, pp. 114-120.

[7] TANG, M., ZHANG, X., LIU, Y., JI, P., FAN,L., TAN,Q., LIANG, T.: *Design and preparation of a high immunity piezo resistive sensor*, Journal of Measurement Science and Instrumentation, Vol.17 No.1 Mar. 2026, doi: 10.62756/jmsi. 1674-8042.2026010.

[8] XUEJIAN, L.: *Design of Safe C language for program verification*, IEEE Access, doi: 10.1109/ACCESS.2026.3656523.

[9] ZHAO,P., ZHANG,Y., CHEN,Y., JIA,Y.: *Modular Programming in C for MCU,* doi: 10.1109/EIECS66708.2025.1128, 2025 5th International Conference on Electronic Information Engineering and Computer Science

[10] KERNIGHAN,B.W., RITCHIE,D.M.: *The C programming language*, Second Edition, Prentice Hall Software Series.

[11] VLATACOM, I.: *159_5_1.3D_Fuze vAF-M17 and Initiator vFI-M17 PCB Testing,* (Internal), 2025, unpublished.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MONTE CARLO BASED PRECISION ANALYSIS OF A LONG-RANGE ARTILLERY PROJECTILE

DAMIR D. JERKOVIĆ
Military Academy, University of Defence, Belgrade, damir.jerkovic@va.mod.gov.rs

NEBOJŠA HRISTOV
Military Academy, University of Defence, Belgrade, nebojsahristov@gmail.com

ALEKSANDRA OBRADOVIĆ
Military Technical Institute, Belgrade, aleksandraobradovic.va@gmail.com

NIKOLA MAČAR
Serbian Armed Forces, Belgrade, macarnikola71@gmail.com

BRANISLAV NIKOLIĆ
Military Academy, University of Defence, Belgrade, bane.nnv@gmail.com

Abstract: *The paper presents a preliminary assessment of the precision measures on the conceptual design of a long-range artillery projectile. The precision measures were calculated using the Monte Carlo method based on the projectile trajectory elements obtained using the flight model. The precision testing procedure is based on numerous simulations of projectile flight with six degrees of freedom, adjusted with selected perturbations of the initial firing conditions. Monte Carlo simulations consisted of 200 virtual firings with standard deviations of initial conditions. The initial conditions consisted of the standard deviation of the: muzzle velocity for different propellant charges, mass of the projectile, initial elevation and azimuth angles and estimated values of the angular velocities around transverse axes. The results of the precision expressed through probable deviations showed good agreement with referenced data of related artillery systems.*

Keywords: *Monte Carlo Method, Long-Range Artillery Projectile, Precision, Probable Deviation, Standard Deviation.*

1. INTRODUCTION

Contemporary artillery systems demand continuous evolution due to extend range, maintain and improve precision and accuracy and enhance operational effectiveness [1, 2]. The global trend is the standardization of mail caliber systems for tactical artillery support, particularly the 155 mm caliber, which has become widely adopted across many countries due to its favorable balance between range, payload, and system compatibility [1]. The artillery system 152 mm as artillery system with similar ballistics characteristics, also depends on technological and standardization design progress.

The research investigates the precision of the conceptual design of velocity enhanced long-range artillery projectile (V-LAP) 155 mm, as part of the ballistic performance.

The objective of this work is to provide a solid assessment of flight performance and precision to identify concurrent design directions for novel long-range artillery projectiles, based on reliable methods.

2. METHODOLOGY

The research consists of the flight model, Monte Carlo model for assessment of the deviations and determination of the precision measures, performed on the conceptual design of V-LAP 155 mm.

2.1. Conceptual Design of Projectile

The final phase of the projectile trajectory analysis is the determination of the precision and efficiency metrics of the projectile model using Monte Carlo simulations [3,4,5], Fig. 1. Due to determination of the trajectory parameters, it is necessary to employ the flight model.

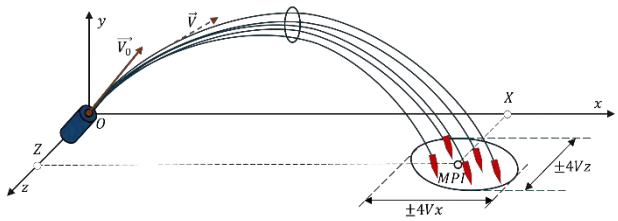


Figure 1. Exterior Ballistics Range Firing

This procedure simulates exterior ballistic range testing, during which the trajectory elements of an unguided projectile are measured [3]. Exterior ballistics firing in these tests varies depending on the type of equipment and the objective of the testing. Range firing is conducted

Corresponding author: D. Jerković;
E-mail: damir.jerkovic@va.mod.gov.rs
DOI: 10.5937/OTEH26023J

using high-explosive projectiles, like the one proposed in this model; their impact can then be monitored and recorded from a safe distance [3,4,5].

The study was performed on the conceptual design of long-range projectile, developed to improve the maximal ranges for artillery system with top propellant charges, Fig. 2.

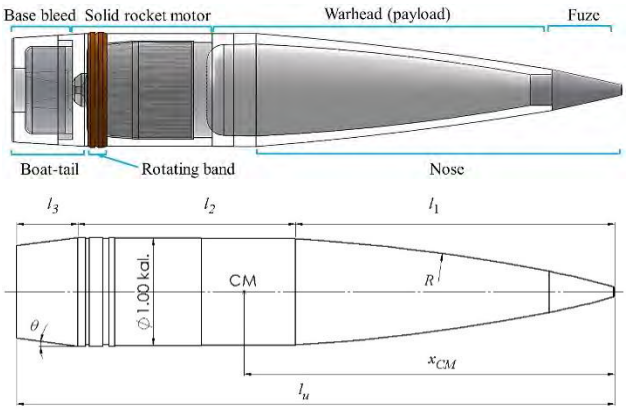


Figure 2. Conceptual design of long-range projectile

2.2. Flight Model

Flight model incorporates model of six degrees of freedom (6-DOF) [3,4,6] to determine projectile trajectory elements and the deviations of the elements.

Governing equations of the 6-DOF model:

$$\frac{d\overrightarrow{r_{CG}}}{dt} = \overrightarrow{V_r} \tag{1}$$

$$\frac{d(m\overrightarrow{V_r})}{dt} = \vec{R} + m\vec{g} + m\overrightarrow{a_c} \tag{2}$$

$$\frac{d(I\vec{\Omega})}{dt} = \overrightarrow{M_R} \tag{3}$$

$$\vec{\Omega} = \overrightarrow{\Omega_E} + \dot{\vec{\lambda}} + \dot{\vec{\varphi}} + \dot{\vec{\psi}} + \dot{\vec{\theta}} + \dot{\vec{\phi}} \tag{4}$$

Equations are applied on the model of long-range projectile, Fig. 1, with determined aerodynamic properties, [7, 8, 9,10].

Parameters in governing equations are:

$\overrightarrow{r_{CG}}$, vector of the center of gravity of projectile,

m, mass of projectile,

I, inertia tensor of projectile,

$\overrightarrow{V_r}$, velocity of the center of gravity of projectile,

$\vec{\Omega}$, angular velocity of projectile,

$\overrightarrow{\Omega_E}$, angular velocity of Earth coordinate frame,

λ, φ, geographic longitude and latitude,

ψ, θ, ϕ, angles of projectile,

$m\vec{g}$, gravity force,

$m\vec{a_c}$, Coriolis force,

$\vec{R}$, aerodynamic force on projectile and

$\overrightarrow{M_R}$, aerodynamic moment on projectile.

2.3. Precision measures

The standard deviation (σ), [11, 12] is a summary measure of the differences of each observation from the mean, i.e. each shot from the mean point of impact (MPI).

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{3}$$

$$\sigma_z = \sqrt{\frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n-1}} \tag{4}$$

Based on the observer's position, it is possible to determine the coordinates of the projectile impact point (X, Z), and then, relative to the position of the firing asset and the direction of fire, calculate the probable deviation of the projectile in range and cross range (Vx, Vz) [3,4,5]. Fig. 1 shows an example of range firing [3].

Parameters Vx and Vz represent the limits of the interval containing 50% of the hits and are determined based on the standard deviation σ_x and σ_z as follows [3,4,5]:

$$V_x = 0.6745\,\sigma_x = 0.6745\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{5}$$

$$V_z = 0.6745\,\sigma_z = 0.6745\sqrt{\frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n-1}} \tag{6}$$

Where are (x_i, z_i), coordinates of the i-th hit, ($\bar{x}$, $\bar{z}$), coordinates of the mean point of impact (MPI), and n is number of hits.

Circular error probable (CEP) [9-12] or circle of equal probability is measure of the precision that represents the radius of the circle, centered on the MPI, such that 50% of the impact points lie within it, Fig. 3. CEP is based on circular bivariate normal distribution (CBN), analog to normal (Gaussian) distribution.

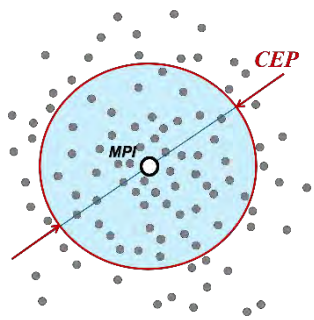


Figure 3. CEP

$$CEP = \begin{cases} \sigma_L(0{,}67 + 0{,}8w^2), & 0 < w < 0{,}5 \\ 0{,}59\sigma_L(1 + w), & 0{,}5 \le w \le 1 \end{cases} \tag{7}$$

$$\sigma_L = max(\sigma_x, \sigma_y), \sigma_S = min(\sigma_x, \sigma_y), w = \frac{\sigma_S}{\sigma_L} \quad (8)$$

3. RESULTS

Monte Carlo simulations consist of applying disturbances to the initial firing conditions, as defined in Table 1. These include the muzzle velocity, azimuth and elevation angle of the barrel, the initial mass of the projectile, and the angular velocities q and r. Each Monte Carlo simulation comprised 200 firings, with pre-defined standard deviations for the analyzed disturbances. As a result, the disturbances varied randomly from one firing to another according to normal (Gaussian) distributions. The standard deviations considered were assumed based on available literature data for similar 155 mm projectiles [13,14].

Table 1. Initial conditions and standard deviations for the Monte Carlo simulation

Initial Condition		Unit	Mean value	Standard deviation
Muzzle velocity	V_{01}	[m/s]	830	2
	V_{02}		950	
Initial mass	m_0	[kg]	43.3	0.1
Elevation angle (barrel elevation)	θ_0	[mils]	996	1
Yaw angle (barrel azimuth)	ψ_0		0	1.25
Pitch angular velocity	q	[s^{-1}]	0	1
Yaw angular velocity	r		0	1

In real firing conditions, disturbances in initial velocity and projectile mass are generally considered systematic deviations. These are corrected using adjustments provided in firing tables. For example, the deviation given in the tables for initial velocity is ±10 m/s, while for the projectile mass, the deviation is related to the number of markings on the ammunition (the so-called weight zones). These systematic deviations also include the following initial conditions: wind speed (longitudinal and lateral), air temperature, atmospheric pressure, and propellant charge temperature. However, the measurement of these quantities during ballistic firing is also subject to random deviations due to the measurement uncertainties of the instruments used. Therefore, the Monte Carlo simulations in this analysis also include disturbances in initial velocity and projectile mass. Disturbances in the azimuth and elevation angle of the barrel are also considered random deviations, as they arise from mechanical factors (vibrations, recoil, etc.) or improper operator handling. Finally, disturbances in the angular velocities q and r are viewed as dynamic, random deviations that cannot be controlled and vary from one firing to another.

In this part of the study, the precision assessment was done to obtain a result of the probable deviation for hybrid model of long-range artillery projectile 155 mm. The initial set of disturbances was performed to muzzle velocity, mass, initial angles of the barrel and initial angular velocities.

4.1. Results of simulation for probable deviation

The Monte Carlo simulation results are shown in Figs. 4 through 13 for an elevation angle of 56° and muzzle velocities V_{01} (830 m/s) and V_{02} (950 m/s), corresponding to maximum ranges of 42 km and 57 km, respectively. Five firing scenarios were considered for each muzzle velocity. Muzzle velocity disturbances were accounted for in the first scenario (Figs. 4 and 6), while projectile mass disturbances were included in the second (Figs. 6 and 7).

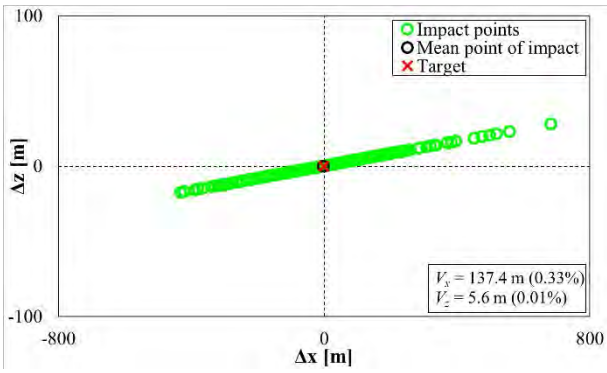


Figure 4. Probable deviation of hits due to deviation of muzzle velocity at 830 m/s

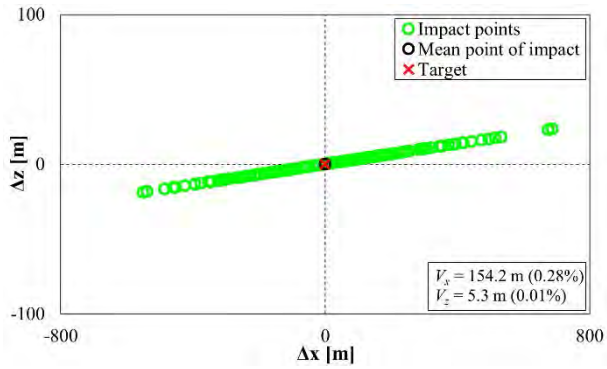


Figure 4. Probable deviation of hits due to deviation of muzzle velocity at 950 m/s

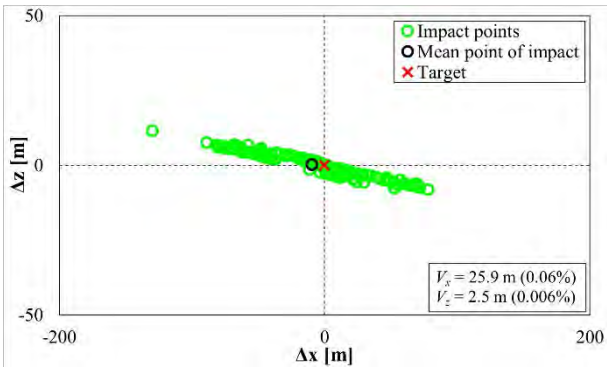


Figure 6. Probable deviation of hits due to deviation of projectile mass at 830 m/s

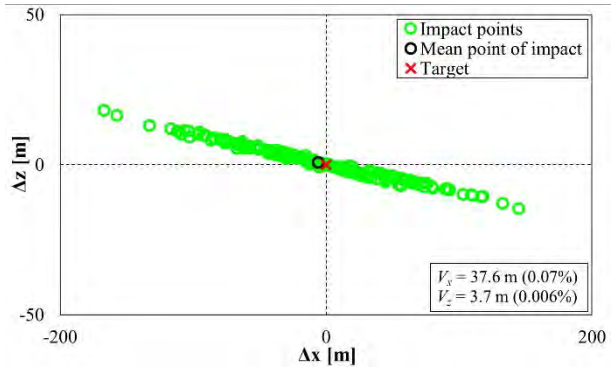


Figure 7. Probable deviation of hits due to deviation of projectile mass at 950 m/s

Azimuth and barrel elevation angle disturbances were examined together in the third scenario (Figs. 8 and 9).

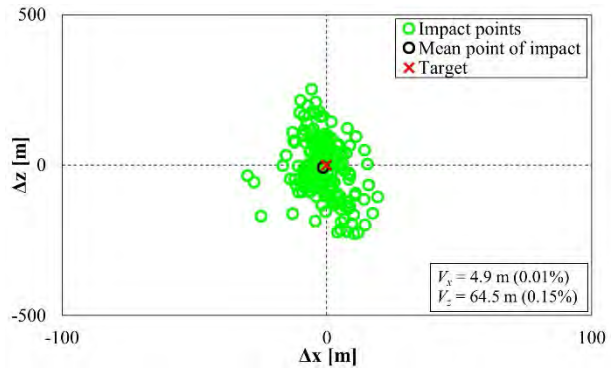


Figure 8. Probable deviation of hits due to deviation of elevation and yaw angles at 830 m/s

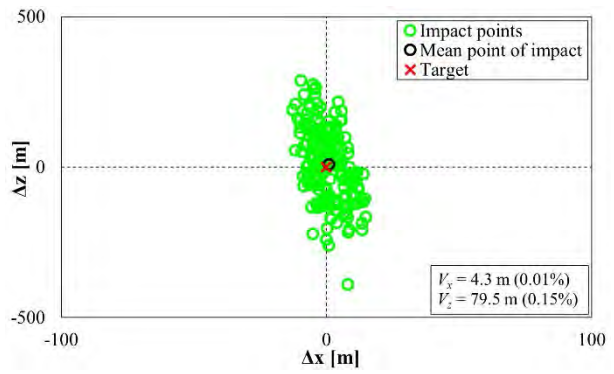


Figure 9. Probable deviation of hits due to deviation of elevation and yaw angles at 950 m/s

Similarly, disturbances in the angular velocities *q* and *r* were introduced in the fourth scenario (Figs. 10 and 11).

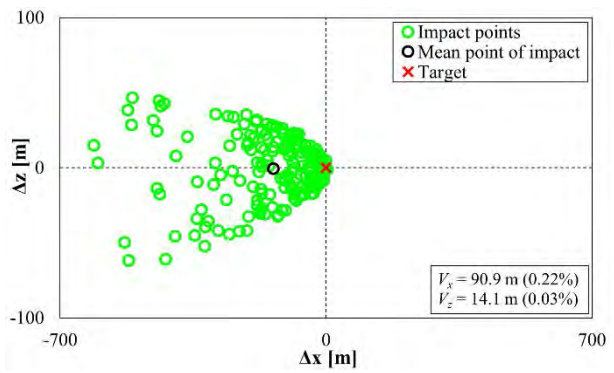


Figure 10. Probable deviation of hits due to deviation of pitch and yaw angular velocities at 830 m/s

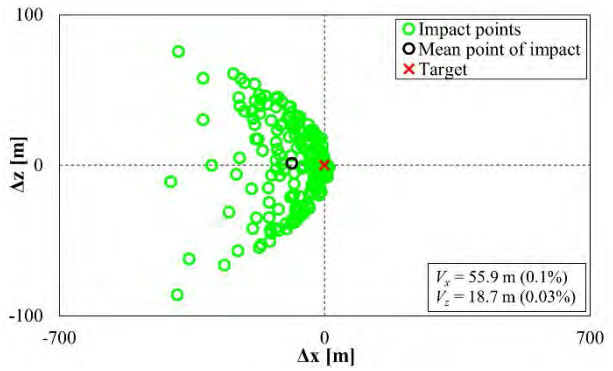


Figure 11. Probable deviation of hits due to deviation of pitch and yaw angular velocities at 950 m/s

Finally, the fifth scenario includes disturbances of all initial conditions considered in the first four scenarios (Figs. 12 and 13).

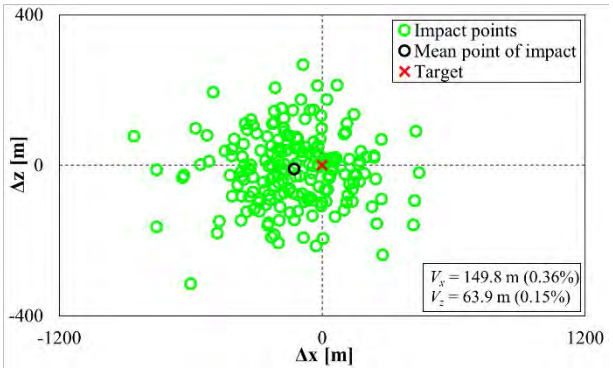


Figure 12. Probable deviation of hits due to deviation of all considered integral factors at 830 m/s

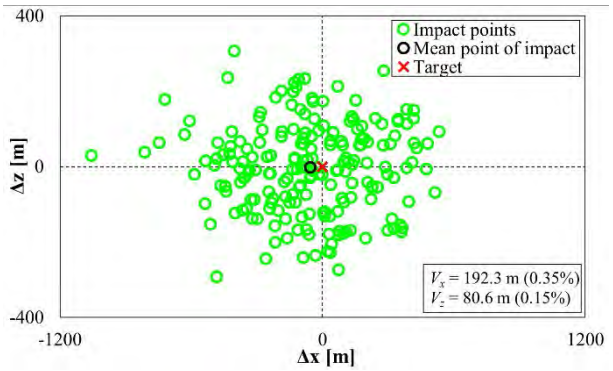


Figure 13. Probable deviation of hits due to deviation of all considered integral factors at 950 m/s

The target is located at point (0, 0) in each figure, and the projectile moves in the positive direction of the *x*-axis, from left to right. The probable deviation of the projectile across the range Vx and deflection Vz for each scenario is also illustrated in the figures. The percentages are calculated relative to the mean range.

The results in Figs. 4, 5, 6 and 7 show that disturbances of muzzle velocity and projectile mass have a significantly greater impact on range than on projectile deflection, with the mean points of impact nearly all centered on the target.

The results of the third scenario in Figs. 8 and 9 demonstrate that disturbances of the elevation angle cause minimal changes in range, whereas disturbances in the

barrel azimuth angle significantly affect projectile deflection. Although the considered standard deviations for both angles are nearly identical, their effects differ substantially. Furthermore, for this scenario, the mean point of impact coincides with the target point.

The hit plot for the fourth scenario shows that disturbances in the angular velocities q and r affect both the range and the projectile deflection (Figs. 10 and 11). All hits are located at distances less than or equal to the target range. The distribution of hits is nearly symmetrical with respect to the x-axis, with a boundary in the shape of an arc directed toward the target, which corresponds to the zone of highest hit density. The mean point of impact deviates from the target point, particularly for the initial velocity V_{01} (Fig. 10).

4.2. Results of simulation for CEP

The dimensions of the CEP metrics are presented in Table 3 under all disturbances.

Table 3. Results of CEP

Velocity [m/s]	CEP [m]	Vx [m]	Vz [m]
830	121.8	149.8	63.9
950	156.3	192.3	80.6

4.3. Analysis

The impact of all considered disturbances on the hit pattern is shown in Fig. 1. In the case of muzzle velocity V_{01}, the hit density is higher in the zone in front of the target, which is why the mean point of impact is at a greater distance from the target compared to the one obtained with velocity V_{02}, which is very close to the target. The hits for velocity V_{02} are more evenly distributed around the target than those associated with velocity V_{01}. This is a direct result of the stronger influence of disturbances in the angular velocities q and r in the case of velocity V_{01}. The probable deviation, as a measure of precision, amounts to 0.35% in range (Vx) and 0.15% in deflection (Vz) relative to the mean range for both velocity cases V_{01} and V_{02}, meaning that the projectile maintains the same precision even at a longer range. Based on the study presented in [9], which refers to 155 mm artillery system with 39 caliber barrel length and 52 caliber barrel length, the probable deviations Vx and Vz should not exceed 0.75% and 0.14%, respectively. Available data [15] on similar existing projectile models show that the probable error of the hybrid V-LAP projectile, which is currently in use, is 0.6% in range (Vx) and 0.13% in deflection (Vz) relative to the firing range. The probable error must remain within these limits to enable the successful application of potential 1D and 2D Course Correction Fuzes [4].

Available literature data show that the proposed projectile model has satisfactory precision, and even better in range. However, this comparative analysis is more indicative than fully objective, as the Monte Carlo simulations performed do not include all possible disturbances that can occur in real firing conditions. Furthermore, the standard deviations used were taken from similar studies available in the literature [13,14], which state that they are appropriate and representative for the analyzed model. However, it is not explained how these deviations were determined, nor are reference studies cited that would confirm this.

5. CONCLUSION

The study presents an investigation into ballistic performance of the long-rang artillery projectile 155 mm through precision assessment. The analysis showed that the considered disturbances of the initial conditions proportionally affect the trajectory parameters. The external ballistics model confirmed that the proposed projectile concept can achieve a substantial increase in range.

The results also demonstrated that range growth can be further improved through complementary approaches, including higher muzzle velocity, aerodynamic optimization, base-drag reduction, and rocket propulsion. The proposed approach provides a solid basis for future design optimization and experimental validation, particularly in relation to precision, flight performance, and range extension capability.

Acknowledgment

This research has been supported by the University of Defence in Belgrade within Project No. VA/TT/1/24-26.

References

[1] Gouveia H. and Freitas R., *Innovations and trends in field artillery weapon systems*, Cogent Social Sciences, 10(2024), 1.

[2] Jelić, Z., *Optimization of design parameters for modular range enhanced projectile*, Ph. D. thesis, Cranfield University, Cranfield, UK, 2015.

[3] D. Regodić, *Exterior ballistics* [in Serbian language], Military Academy, Belgrade, Serbia, 2006.

[4] Regodić D., *Solved problems collection in external ballistics* [in Serbian language], Military Academy, University of Defence, Belgrade, Serbia, 2003.

[5] I. Todić, D. Ćuk, S. Pajić, Flight Dynamics of Projectiles, Faculty of Mechanical Engineering, University of Belgrade, Belgrade, Serbia, 2021.

[6] McCoy R. L., *Modern exterior ballistics*, Schiffer Publishing Ltd, 1998.

[7] Ferfouri A., Allouche T., Jerković D., Hristov N., Vučković M., and Benmeddah A., *Prediction of drag aerodynamic coefficient of the 155mm projectile under axisymmetric flow using different approaches*, Journal of the Serbian Society for Computational Mechanics, 17(2023), 2, pp. 69–86.

[8] Ferfouri A., Jerković D., Kari A., and Živković S., *Computational analysis of rocket-assisted projectile trajectory*, 11th International Scientific Conference on Defensive Technologies, Belgrade, Serbia, 2024, pp. 15–23.

[9] Ferfouri, A., Jerković, D., Hristov, N., Kari, A. V,

and Allouche, T., *Performance analysis of grid topologies and rans turbulence models in predicting aerodynamic drag coefficient at zero-yaw for an artillery projectile*, Journal of Applied Fluid Mechanics, 18(2025), 3, pp. 585–600.

[10] Ferfouri, A., Jerković, D., Živković S., Allouche T., and Hristov N., *Structural Integrity Assessment of a Long-Range Artillery Projectile Body via Analytical and Numerical Approaches*, Journal of Applied and Computational Mechanics, 2025.

[11] Uroš, A., Danica, S., Slavica, T., Ivan, D., & Zoran, B. (2018). Improvement of experimental methods for detonation of velocity determination, International Scientific Conference on Defensive Technologies, OTEH (ID082). Belgrade, Republic of Serbia, Ministry of Defence, Military Technical Institute Belgrade.

[12] Morris, D. R. (2013) Weaponeering: Conventional Weapon System Effectiveness, 2nd Edition. Reston, Virginia 20191-4344, American Institute of Aeronautics and Astronautics, Inc.

[13] A. Szklarski, R. Głębocki, M. Jacewicz, Impact point prediction guidance parametric study for 155 mm rocket assisted artillery projectile with lateral thrusters, Archive of Mechanical Engineering (2020) 31–56. https://doi.org/10.24425/ame.2020.131682

[14] N. Hamel, E. Gagnon, CFD and Parametric study on a 155 mm artillery shell equipped with a roll-decoupled course correction fuze, in: 29th AIAA Applied Aerodynamics Conference, American Institute of Aeronautics and Astronautics, Reston, Virigina, 2011. https://doi.org/10.2514/6.2011-3027

[15] Rheinmetall Denel Munition, ASSEGAI Projectile Family, (n.d.). https://www.valka.cz/attachments/3583/Assegai_Projectile_Family__2021_.pdf (accessed February 12, 2025)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SHEAR STRESS-BASED FORCE MODELING FOR M-84 TRACKED VEHICLE ON DEFORMABLE TERRAIN

MOMIR DRAKULIĆ
University of Defence in Belgrade, Military Academy, Belgrade, momir.drakulic@va.mod.gov.rs

ALEKSANDAR ĐURIĆ
University of Defence in Belgrade, Military Academy, Belgrade, aleksandar.djuric@va.mod.gov.rs

Abstract: *This paper presents a methodology for calculating interaction forces acting on the tracked armoured vehicle M-84 during motion on deformable terrain. The research is based on a previously developed and experimentally validated track-terrain interaction model, which is extended toward application at the vehicle level. The proposed approach applies a shear stress model within the framework of terramechanics theory to estimate the distribution of stresses along the track-soil contact area. Based on this model, the resultant interaction forces are numerically evaluated under different operating conditions, including straight motion and steering. The calculated force outputs are structured for integration into a complete dynamic vehicle model. They also provide a consistent input set for driveline analysis and transmission studies, especially for optimization of the gear shifting process in a power-shift transmission system, aiming to achieve smooth gear transitions without interruption of power flow. The presented methodology contributes to improved prediction of mobility performance and supports further development of control-oriented models of tracked armoured vehicles.*

Keywords: *Tracked vehicle, terramechanics, shear stress model, track-soil interaction, vehicle dynamics.*

1. INTRODUCTION

Tracked vehicles are designed to maintain mobility across terrains that vary significantly in stiffness, strength, and surface conditions [1,2]. Their operational performance depends not only on propulsion and suspension systems but also on the complex mechanics of track–terrain interaction [3,4]. During high-speed steering manoeuvres, nonlinear distributions of contact forces emerge, which strongly affect vehicle stability, controllability, and energy consumption [5]. These effects become especially pronounced in heavy platforms such as main battle tanks, where large inertial loads amplify slip, lateral displacement, and uneven load distribution along the track [6].

Accurate prediction of track–terrain interaction forces is essential for reliable assessment of vehicle behaviour under demanding operating conditions [7]. Empirical models provide simplified estimates but often fail to capture the coupled effects of soil deformation and slip dynamics [8]. On the other hand, high-fidelity numerical methods, such as discrete element or finite element approaches, offer improved physical realism but are typically computationally expensive for system-level simulations. Therefore, a balanced modelling approach that combines physical consistency with computational efficiency is required for engineering analysis and design purposes.

The simulation framework used in this study is based on terramechanics principles and the shear stress theory of soil deformation, enabling the prediction of longitudinal and lateral contact forces generated during track–terrain interaction [9]. The model has previously been developed and validated for the infantry fighting vehicle BVP M80, with results reported in [10], where good agreement with experimental data was demonstrated in terms of contact force distribution and vehicle trajectory prediction.

In this work, the same validated modelling framework is applied to the M84 main battle tank. The objective is to investigate how changes in vehicle mass, geometry, and inertial properties influence the predicted track–terrain interaction forces during steering manoeuvres. By introducing vehicle-specific parameters, the study evaluates the adaptability of the model to a heavier and dynamically different tracked platform, while maintaining the same underlying physical formulation.

The contribution of this paper lies in demonstrating the transferability of a previously validated terramechanics-based simulation model to a different class of tracked vehicle and in analysing the resulting differences in contact force distribution under identical maneuver conditions. The results provide additional insight into the influence of vehicle configuration on track–terrain interaction and support the use of such models in comparative studies, system design, and further development of control-oriented simulation tools.

2. MODELING

Considering the variety of approaches used in the analysis of track–terrain interaction, a systematic overview of modern terramechanics methods is given in [8,11]. To achieve a balance between physical realism, numerical stability, and computational efficiency, the Janosi–Hanamoto semi-empirical model is adopted as the basis for

Corresponding author: M. Drakulić;
E-mail: momir.drakulic@va.mod.gov.rs

DOI: 10.5937/OTEH26024D

describing the shear behaviour of deformable terrain. This choice is consistent with the modelling framework previously developed and validated in [10], where the same approach was successfully applied to the infantry fighting vehicle BVP M80.

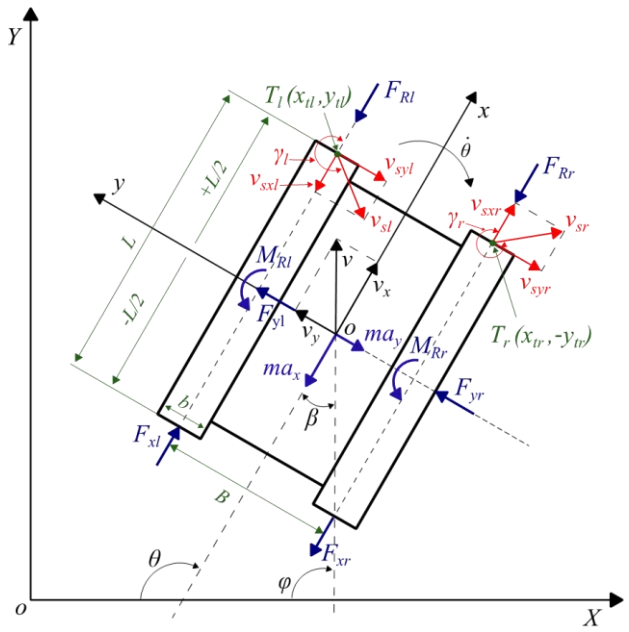


Figure 1. Tracked vehicle during the steering [10].

Starting from the established modelling framework, a consistent vehicle dynamics formulation is required to enable coupling with the terramechanics model [8,12,13]. The vehicle is considered as a rigid body performing planar motion during steering, where the interaction between the left and right track systems generates longitudinal and lateral motion as well as yaw dynamics. A detailed derivation of the governing equations of motion is provided in [10] and is directly utilized in the present study without modification of the underlying structure.

Within the Janosi–Hanamoto formulation, the key coupling variable between vehicle kinematics and soil response is the shear displacement, which arises due to the relative motion between the track and the deformable soil [14,15]. This quantity governs the development of shear stress within the contact patch and provides a direct link between slip conditions and traction generation. The shear stress is described by the exponential relationship:

$$\tau = (c + \sigma tan\phi)\left(1 - e^{-\frac{j}{K}}\right) \quad (1)$$

where c is the soil cohesion, σ is the normal stress, φ is the internal friction angle, j is the shear displacement, and K is the shear deformation modulus.

Based on the obtained shear stress distribution, the interaction forces between the track and the terrain are determined by integrating the local stresses over the contact length. In this way, the longitudinal and lateral force components acting on the right and left track are obtained as functions of vehicle motion, soil properties, and local pressure distribution [16–18].

Based on Figure 1, F_{xr} and F_{xl} represent the components of the shear force exerted by the right and left tracks, respectively, on the terrain in the x-direction. F_{yr} and F_{yl} are the corresponding components of the shear force in the y-direction. M_{Rr} and M_{Rl} denote the steering resistance torques for the right and left tracks, respectively, while F_{Rr} and F_{Rl} represent the rolling resistances for the tracks on both sides.

$$F_{xr} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_r \tan\phi)\left(1 - e^{-\frac{j_r}{K}}\right) b\cos(\gamma_r + \pi)dx \quad (2)$$

$$F_{xl} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_l tan\phi)\left(1 - e^{-\frac{j_l}{K}}\right) b\cos(\gamma_l + \pi)dx \quad (3)$$

$$F_{yr} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_r tan\phi)\left(1 - e^{-\frac{j_r}{K}}\right) b\sin(\gamma_r + \pi)dx \quad (4)$$

$$F_{yl} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_l tan\phi)\left(1 - e^{-\frac{j_l}{K}}\right) b\sin(\gamma_l + \pi)dx \quad (5)$$

In addition to traction forces, the model accounts for rolling resistance effects through a load-dependent formulation that includes lateral load transfer during steering manoeuvres. The normal load distribution between the left and right tracks is influenced by vehicle mass, lateral acceleration, centre of mass height, and track gauge, which leads to asymmetric resistance forces on the two sides of the vehicle [19]. This effect is incorporated in the form of equivalent resistance forces acting at the track level, ensuring compatibility with the overall dynamic equilibrium.

$$F_{Rr} = f\left(\frac{mg}{2} + \frac{ma_y h}{B}\right) \quad (6)$$

$$F_{Rl} = f\left(\frac{mg}{2} - \frac{ma_y h}{B}\right) \quad (7)$$

$$M_{Rr} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_r tan\phi)\left(1 - e^{-\frac{j_r}{K}}\right) b\sin(\gamma_r + \pi)xdx \quad (8)$$

$$M_{Rl} = \int_{-\frac{L}{2}}^{\frac{L}{2}} (c + p_l tan\phi)\left(1 - e^{-\frac{j_l}{K}}\right) b\sin(\gamma_l + \pi)xdx \quad (9)$$

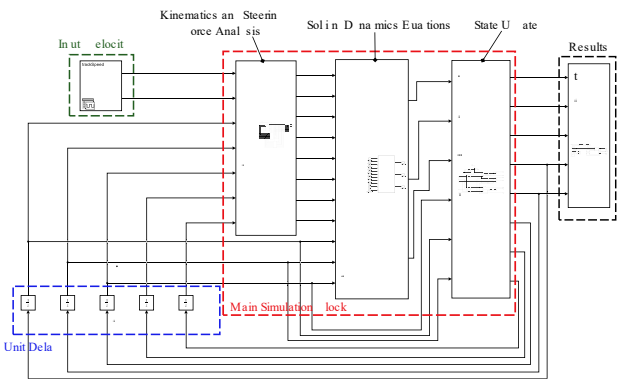


Figure 2. Simulation model [10].

The developed force and moment model is implemented in a MATLAB/Simulink environment, where it is coupled with the previously established vehicle dynamics model. The simulation structure enables numerical evaluation of high-speed steering manoeuvres by computing

instantaneous track–terrain interaction forces and vehicle response in a time-domain framework. Compared to the previously published model in [10], the present implementation retains the same mathematical structure while extending its application to a different vehicle class by introducing the geometric and inertial parameters of the M84 main battle tank, shown in table 1.

Table 1. Vehicle Parameters.

P me e	*V e*	*Uni*
m	13 850	[k]
L	3.3	[m]
B	2.526	[m]
b	0.31	[m]
h	1.139	[m]
	0.258	[m]
i	5.786	–

3. RESULTS

For the purposes of simulation, a Gault clay soil was selected as the terrain type (Table 2), as it is representative of cohesive soil conditions commonly associated with the operational environments of the M84 tracked vehicle. The reference trajectory to be followed by the vehicle was predefined [20,21], and a closed-loop control structure based on a PID controller was implemented to ensure accurate trajectory tracking. The controller continuously minimizes the position error between the desired and actual vehicle path, thereby enabling the system to follow the prescribed trajectory while maintaining stable dynamic response throughout the maneuver. The simulated trajectory, shown in Figure 3, closely follows the prescribed reference path throughout the entire manoeuvre.

Table 2. Terrain Parameters.

P me e	*G*	*Uni*
c	68.92	[kPa]
φ	34	[^{0}]
K	0.6	[cm]

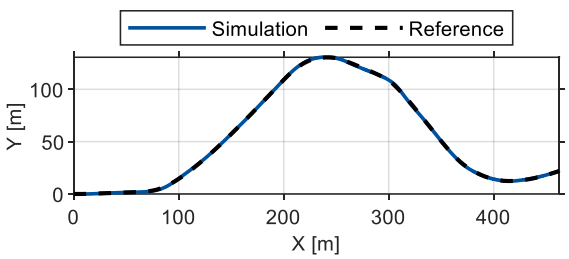


Figure 3. Vehicle trajectory.

This confirms that the PID-based closed-loop control structure maintains adequate path-tracking accuracy despite the combined influence of nonlinear shear-stress traction forces and lateral load transfer acting on the vehicle.

The corresponding sideslip angle, presented in Figure 4, remains within a narrow ran e of a roximatel −0.4° to +0.2°, with the most pronounced excursions coinciding with the steepest curvature segments of the trajectory (aroun t ≈ 10 s an t ≈ 40–55 s). This behaviour confirms the expected coupling between path curvature and lateral slip and indicates that lateral instability effects remain limited for the considered manoeuvre and terrain conditions.

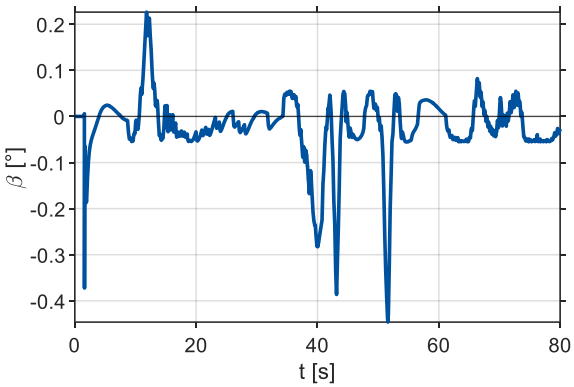


Figure 4. Sideslip angle.

This kinematic response is governed by the underlying track–terrain interaction forces. The longitudinal forces on the right and left track, shown in Figures 5 and 6 respectively.

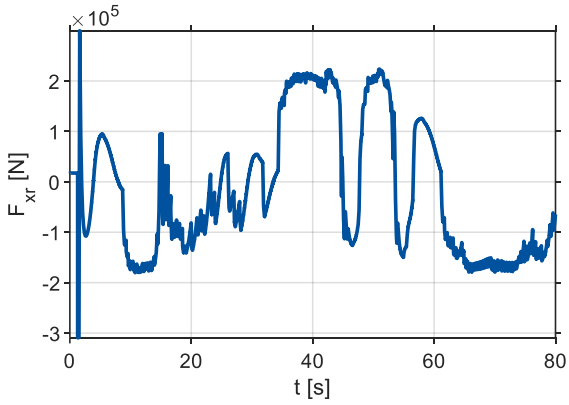


Figure 5. Longitudinal force on the right track.

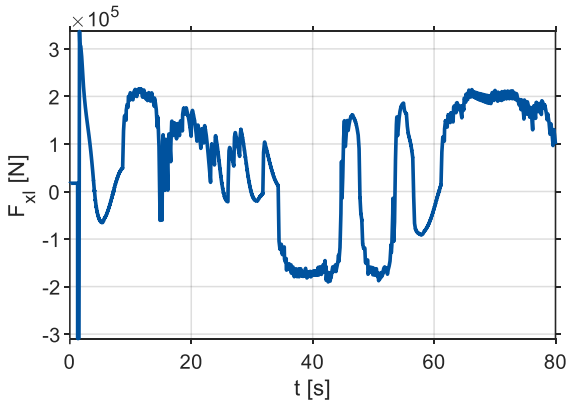


Figure 6. Longitudinal force on the left track.

This asymmetry between Fxr and Fxl is consistent with the

differential traction and braking distribution required for skid-steering and represents the primary mechanism generating the yaw moment needed to execute the prescribed turn; the pronounced plateau visible between t ≈ 30–55 s corresponds to the sustained turning phase, during which the outer track is required to sustain a higher traction level than the inner track.

A similar coupling is observed in the lateral force components, presented in Figures 7 and 8, which remain close to zero during straight-line motion and develop pronounced peaks of up to approximately 3.5×10^4 N urin the steering phase.

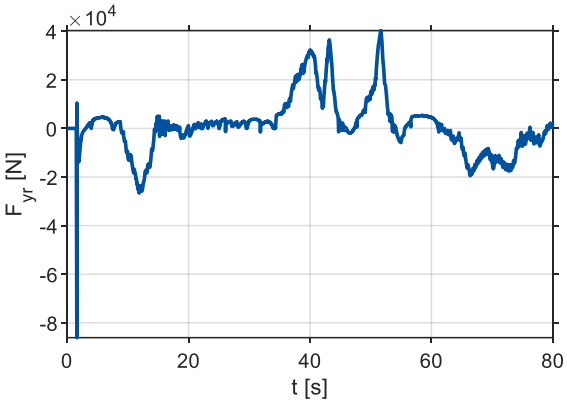


Figure 7. Lateral force on the right track.

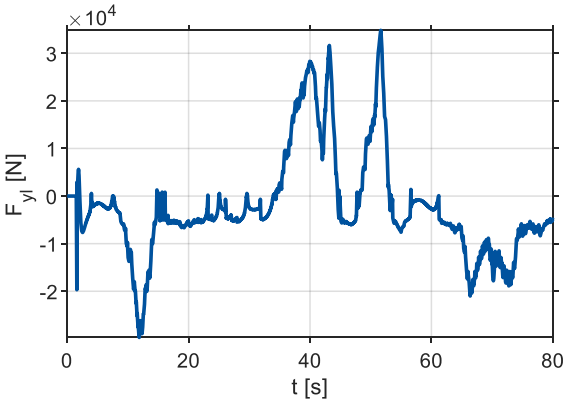


Figure 8. Lateral force on the left track.

The magnitude and timing of these lateral force peaks correlate well with the sideslip angle response discussed above, confirming the physical consistency of the shear-stress-based lateral force formulation across both tracks. The combined effect of these longitudinal and lateral force asymmetries is reflected in the turning resistance moment shown in Figure 9, which varies between approximately -5×10^5 Nm an 5×10^5 Nm and reaches its largest sustained values during the central turning segment of the manoeuvre; the relatively high magnitude of this moment, compared with the values previously reported for the lighter BVP M80 platform in [10], illustrates the influence of the M-84's larger mass and track gauge on rolling and turning resistance.

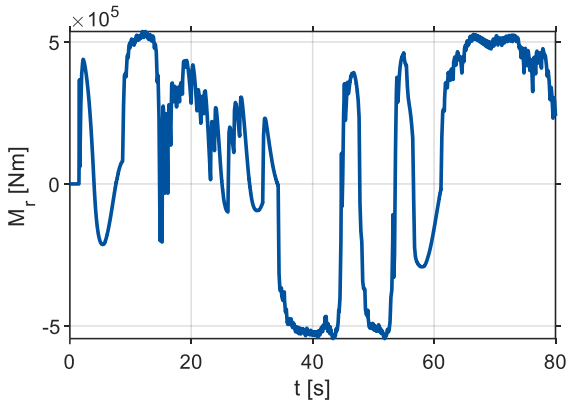


Figure 9. Turning resistance moment.

Despite these substantial variations in track forces and moments, the vehicle's speed response remains well-regulated. Following an initial transient associated with the start of the manoeuvre, the total vehicle speed, shown in Figure 10, stabilises around 25 km/h with only minor fluctuations of approximately ±1 km/h during the steering phase.

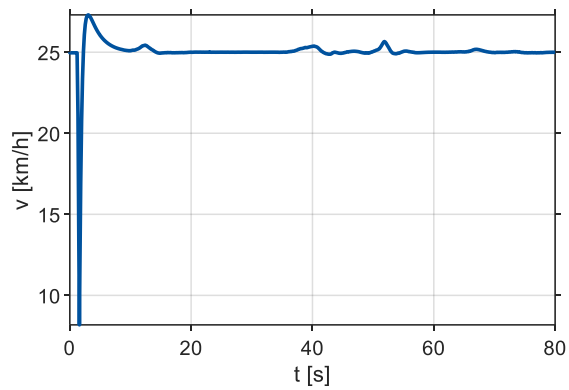


Figure 10. Total vehicle speed.

The longitudinal velocity component, presented in Figure 11, closely tracks this overall speed profile, confirming that the lateral velocity contribution remains small throughout the manoeuvre.

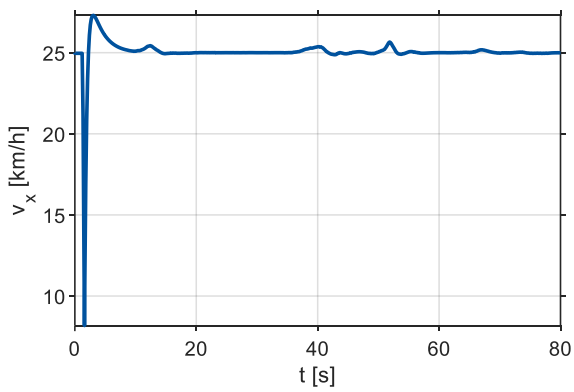


Figure 11. Longitudinal vehicle speed.

This is corroborated by the lateral velocity response in Figure 12, which stays within a narrow band of approximately ±0.15 km/h; although small in absolute terms, the visible oscillations during the steering phase are

directly linked to the lateral force and sideslip behaviour discussed above and represent the residual lateral dynamics not fully suppressed by the trajectory controller.

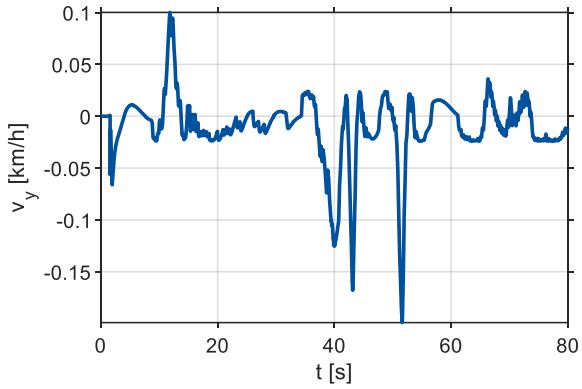


Figure 12. Lateral vehicle speed.

The corresponding acceleration responses further support this picture of stable, well-controlled vehicle behaviour. Apart from an initial transient peak of approximately 1.5 m/s² associated with the vehicle's start-up phase, the longitudinal acceleration in Figure 13 remains close to zero, with only minor oscillations occurring during the steering segment as the controller works to maintain constant forward speed under varying traction demand.

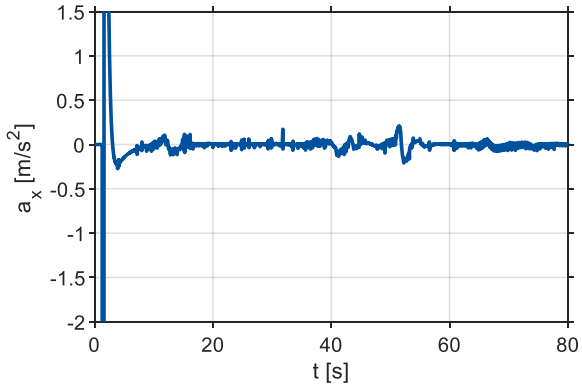


Figure 13. Longitudinal vehicle acceleration.

The lateral acceleration, shown in Figure 14, reaches peak values of approximately ±1.1 m/s² during the sharpest turning segments, in direct correspondence with the lateral force peaks discussed for Figures 7 and 8, these values remain within a range that does not indicate critical lateral instability, further supporting the conclusion that the controller and the underlying force model produce physically consistent vehicle behaviour.

Finally, the trajectory tracking error, presented in Figure 15, oscillates within approximately ±0.6 m, with the largest transient excursions occurring during the most demanding curvature changes of the manoeu re (t ≈ 15 s, 40 s an 50 s). While these peaks indicate a temporary degradation of tracking accuracy under high lateral force demand, the error consistently returns toward zero, confirming the overall stability and adequacy of the PID control structure for the considered class of high-speed steering manoeuvres.

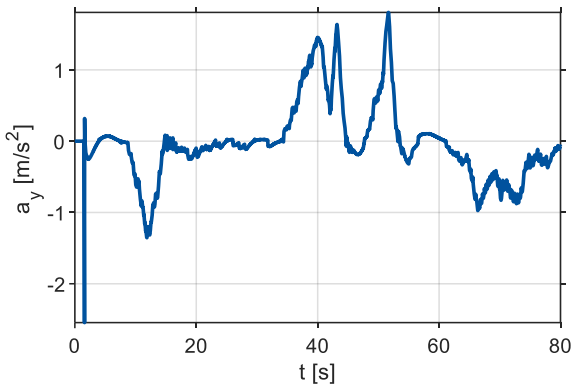


Figure 14. Lateral vehicle acceleration.

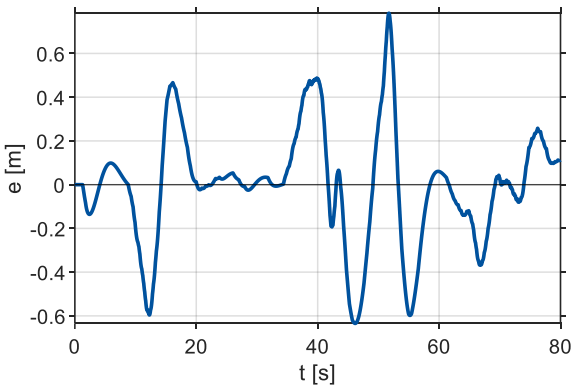


Figure 15. Trajectory tracking error.

4. CONCLUSION

This paper presented a shear stress-based methodology for calculating track–terrain interaction forces acting on the M-84 main battle tank during steering manoeuvres on deformable terrain. The approach extends a previously developed and experimentally validated terramechanics model, originally established for the infantry fighting vehicle BVP M80, to a heavier and dynamically different tracked platform by introducing the M-84's specific geometric and inertial parameters.

The simulation results, obtained for a representative cohesive soil (Gault clay) and a closed-loop PID-controlled steering manoeuvre, demonstrate that the model produces physically consistent and well-correlated force, moment, velocity and acceleration responses. The vehicle maintained accurate trajectory tracking, with sideslip angle and lateral acceleration remaining within bounded, stable ranges throughout the manoeuvre, while the trajectory tracking error remained limited even during the most demanding curvature transitions. The obtained results confirm the model's transferability across vehicle classes while preserving the same underlying physical formulation validated in [10].

The calculated force and moment outputs provide a consistent and structured data set suitable for integration into a complete vehicle dynamics model, and for driveline and transmission studies, including the optimisation of gear-shifting strategies in power-shift transmission

systems. Future work will focus on the experimental validation of the model for the M-84 platform, the extension of the analysis to additional terrain types and manoeuvre conditions, and a direct quantitative comparison between the M-84 and BVP M80 force responses to further characterise the influence of vehicle mass and geometry on track–terrain interaction.

Acknowledgments

This work was supported by the Ministry of Defense, Serbia, under the project „Modeling of Automotive Vehicle Assemblies for the Optimization of Dynamic Parameters", No. A -TT/1/26-28.

References

[1] Wong JY. Theory of ground vehicles. John Wiley & Sons; 2008.

[2] Wong JY. Terramechanics and off-road vehicle engineering: Terrain behaviour, off-road vehicle performance and Design. Elsevier; 2010.

[3] Garber M, Wong JY. Prediction of ground pressure distribution under tracked vehicles—II. Effects of design parameters of the track-suspension system on ground pressure distribution. Journal of Terramechanics 1981;18:71–9. https://doi.org/10.1016/0022-4898(81)90001-X.

[4] Tao J, Deng Z, Cao X, Hu G, Wang P. Modeling and Dynamic Characteristics of Tracked Vehicle Equipped with Symmetrical Suspensions Based on Multi-Body Dynamics and Discrete Element Coupling Method. Applied Sciences 2024;14:10618. https://doi.org/10.3390/app142210618.

[5] Golanbari B, Mardani A, Hosainpour A, Taghavifar H. Predicting terrain deformation patterns in off-road vehicle-soil interactions using TRR algorithm. Journal of Terramechanics 2024;117:101021. https://doi.org/10.1016/j.jterra.2024.101021.

[6] Lv G, Lv H, Chen Z, Chu W, Mao J. Research on inertia transfer in load simulation of tracked vehicle under complex working conditions. Front Mater 2023;10:1188411. https://doi.org/10.3389/fmats.2023.1188411.

[7] Öz emir MN, Kılıç , Ünlüso YS. A new contact & slip model for tracked vehicle transient dynamics on hard ground. Journal of Terramechanics 2017;73:3–23. https://doi.org/10.1016/j.jterra.2017.07.001.

[8] Chakravarthy PN, Sharma PK. Terramechanics Models for Tracked Vehicle Terrain Interaction Analysis A Review. Defence Science Journal 2024;74:469–81. https://doi.org/10.14429/dsj.74.19029.

[9] Janarthanan B, Padmanabhan C, Sujatha C. Lateral dynamics of single unit skid-steered tracked vehicle. International Journal of Automotive Technology 2011;12:865–75. https://doi.org/10.1007/s12239-011-0099-4.

[10] Drakulic M, Peric S, Djuric A, Rakic S. Track–terrain interaction during steering of high-speed tracked vehicles: Numerical modeling and experimental validation. Measurement 2025;256:118225. https://doi.org/10.1016/j.measurement.2025.118225.

[11] Jia W, Liu X, Zhang C, Qiu M, Tan Y, Yu Z. Design of Zero-Differential steering Controller for tracked vehicles with Hydraulic-Mechanical transmission based on particle swarm optimization Algorithm. IEEE Access 2023;11:32187–200. https://doi.org/10.1109/access.2023.3262515.

[12] Gorsich D, Vantsevich V, Paldan J, Moradi L. Modelling and simulation fundamentals in design for ground vehicle mobility Part II: Western approach. Journal of Terramechanics 2024;117:101023. https://doi.org/10.1016/j.jterra.2024.101023.

[13] Vantsevich V, Gorsich D. Modeling and simulation fundamentals in design for ground vehicle mobility part I: Eastern approach. Journal of Terramechanics 2024;117:101020. https://doi.org/10.1016/j.jterra.2024.101020.

[14] Ali Shaikh S, Li Y, Ma Z, Ali Chandio F, Hussain Tunio M, Ahmad F, et al. Effects of soil mechanical properties on the height and tractive performance of rubber grouser at different moisture contents. International Journal of Agricultural and Biological Engineering 2022;15:31–7. https://doi.org/10.25165/j.ijabe.20221506.7137.

[15] Meywerk M, Fortmüller T, Fuhr B, Baß S. Real-time model for simulating a tracked vehicle on deformable soils. Advances in Mechanical Engineering 2016;8:168781401664788. https://doi.org/10.1177/1687814016647889.

[16] Tota A, Galvagno E, Velardocchia M. Analytical study on the cornering behavior of an articulated tracked vehicle. Machines 2021;9:38. https://doi.org/10.3390/machines9020038.

[17] Tota A, Velardocchia M, Rota E, Novara A. Steering behavior of an articulated amphibious All-Terrain tracked vehicle. SAE Technical Papers on CD-ROM/SAE Technical Paper Series 2020. https://doi.org/10.4271/2020-01-0996.

[18] Dimauro L, Venturini S, Tota A, Galvagno E, Velardocchia M. High-speed tracked vehicle model order reduction for static and dynamic simulations. Defence Technology 2024;38:89–110. https://doi.org/10.1016/j.dt.2024.01.006.

[19] Garber M, Wong JY. Prediction of ground pressure distribution under tracked vehicles—I. An analytical method for predicting ground pressure distribution. Journal of Terramechanics 1981;18:1–23. https://doi.org/10.1016/0022-4898(81)90015-x.

[20] Drakulić M, Stanko ić M. ADRC-Based Trajectory Tracking of Unmanned Tracked Vehicles Under High Slippage Disturbance. 2024 14th International Conference on Electrical Engineering (ICEENG), IEEE; 2024, p. 362–6.

[21] Stankovic M, Ting H, Madonski R. From PID to ADRC an back: Ex ressin error-base acti e disturbance rejection control schemes as standard industrial 1DOF and 2DOF controllers. Asian Journal of Control 2024;26:2796–806. https://doi.org/10.1002/asjc.3373.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A NEW METHOD FOR INTERPOLATION OF ROCKET MOTOR THRUST - TIME CURVE FOR ARBITRARY PROPELLANT TEMPERATURE

MIODRAG ĆURČIN
KOL-15N Engineering Belgrade, miodragcurcin@gmail.com

GORAN MARJANOVIĆ
Military Technical Institute, Belgrade, gmarjanovic1961@gmail.com

SRĐAN STOJKOVIĆ
HK Krušik, Valjevo, ssrdjanvti@gmail.com

BOJAN TANASKOVSKI
Ministry of Defence, Belgrade, tanaskovskibojan@yahoo.com

Abstract: *A new algorithm is developed to interpolate the thrust curve for a given propellant temperature based on the known thrust functions for one higher and one lower temperature than the given one. The algorithm is based on the fact that the impulse of thrust as a function of the burnt mass of propellant changes very little depending on the temperature of the propellant compared to the change in thrust and operating time depending on the temperature of the propellant. Examples of calculations for several rocket motors of different purposes are given, which demonstrate the high efficiency of the algorithm.*

Keywords: *solid propellant rocket motor, thrust interpolation, propellant temperature, impulse..*

List of Symbols

F , F_k	thrust at any time and its discrete value [N, daN, kN]
$\bar{F}$	scaled thrust to the mean value of total impulse
I_t	impulse of thrust at time t , [Ns, daNs, kNs]
$\bar{I}_t$	scaled impulse to the mean value of total impulse
I_{tot}	total impulse
$I_{\text{tot}.m}$	mean value of total impulse for plus and minus temperature
T_p	temperature of rocket motor propellant, [°C]
$T_{p.\max}$, $T_{p.\min}$	maximal and minimal temperature for which thrust is known (measured)
t , t_i	time and its discrete value, [s, ms]
t_b	total burning time, [s, ms]
$\bar{t}$	scaled time to the mean value of total impulse
$\bar{\sigma}_m$	coefficient of interpolation quality

1. INTRODUCTION

It is well known that propellant temperature influences operation of solid propellant rocket motor. With increase of propellant temperature level of thrust increases and the operation time decreases, and vice verse - with decrease of propellant temperature level of thrust decreases and the operation time increases [1]. Depending mostly on type of propellant, than on the grain shape and motor construction the influence can vary from small to large amount.

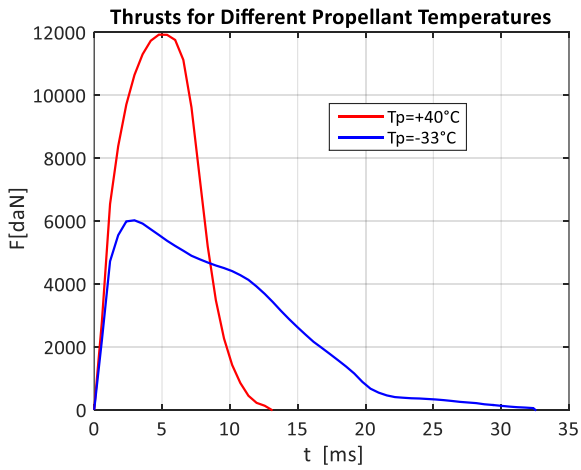


Figure 1. Thrust vs time for two different temperatures for AT rocket M79 OSA.

In figure 1 the thrust for the two propellant temperatures +40°C (red line) and -33°C (blue line) for the anti-tank shoulder launch rocket 90 mm M79 OSA [2] are shown. It can be clearly seen that maximum thrust for the +40°C propellant temperature is two times greater than for -33°C and that operation time for +40°C propellant temperature is two times less than for the -33°C propellant temperature. Note that curves are obtained after filtering and averaging ([3]) of the set of measurements of thrust (seven runs, fig. 2), which is not subject of this work.

Corresponding author: M. Ćurčin;
E-mail: miodragcurcin@gmail.com

DOI: 10.5937/OTEH26025C

In operational usage propellant temperature can be any between minimum and maximum specified for the rocket, and we need thrust function in time for that temperature to calculate trajectory.

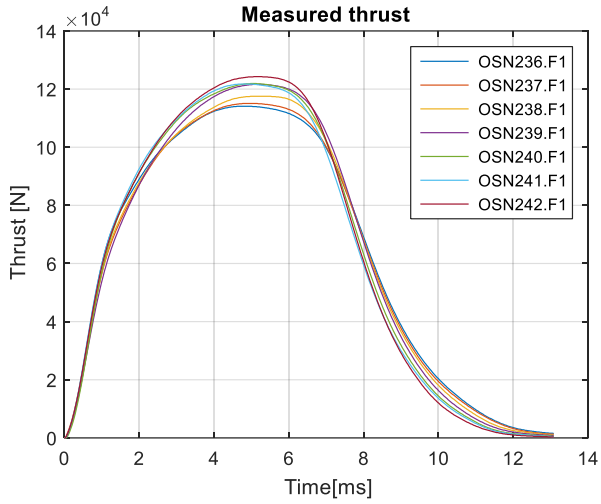


Figure 2. Filtered thrust vs time for 40°C propellant temperature – seven measurements.

The difference of the operation time of rocket motor between the two temperatures can be large while at the same time the difference between the total impulses for the two temperatures are relatively small. For example, for the considered rocket for the +40°C propellant temperature the total impulse is 833 Ns and for the -33°C propellant temperature the value is 795 Ns, the difference being 4.78 %. Total impulse with respect to propellant temperature $I_{\text{tot}}(T_p)$ for rocket motor is shown in figure 3. We can see huge influence at lower propellant temperature and small influence of the propellant temperature at high propellant temperature to the total impulse.

In serial production we usually measure thrust characteristics at extreme temperatures – plus maximum $T_{p.\max}$ and minus maximum operation temperature $T_{p.\min}$, and sometimes also at standard temperature (20°C). But in real combat situation the propellant temperature can be any value in the interval from minimum to maximum and we need the thrust function in time for the actual propellant temperature for many reasons, the main being calculation of trajectory. The question is whether we can determine the thrust in time for any other propellant temperature than the two extremes, for example for $T_p = 20\,°\text{C}$ for the rocket 90 mm M79 OSA based on two known curves as shown in figure 1.

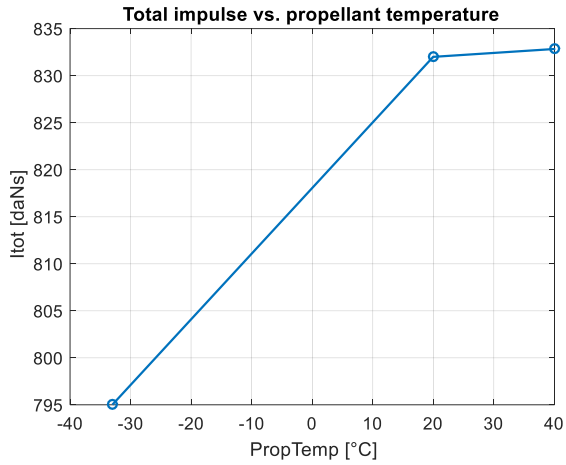


Figure 3. Total impulse with respect to propellant temperature for AT rocket M79 OSA.

Even if we have measured the thrust in time for three or four propellant temperatures the actual propellant may differ from those measured and we have to estimate thrust for the required propellant temperature.

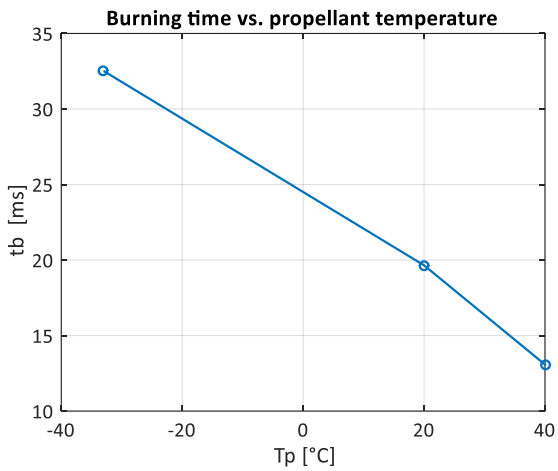


Figure 4. Burning time with respect to propellant temperature for AT rocket M79 OSA.

Statement of the Problem

Suppose that the function thrust with respect to time is known numerically for the two extreme temperature

$$F_i(T_{p.\max}, t_i),\ i = 1,2,..,n_+$$

$$F_j(T_{p.\min}, t_j),\ j = 1,2,..,n_-$$

where n_+ – number of data for plus temperature, and n_- – number of data for minus temperature, and maybe for one or two temperatures between them. Our task is to determine the $F_k(T_p, t_k), k = 1,2,..,m$ at any required temperature T_p, and time instant t_k in m points. One immediately can say "do interpolation between the two curves for known temperature". But the question is how to interpolate between two curves which do not have neither same abscissa nor same ordinate? Let us consider thrust functions for the rocket OSA (figure 1). The curves for maximum propellant temperature (40°C) and minimum temperature (-33°C) have totally different span in abscissa

and ordinate and direct interpolation between them is not possible. Suppose that we want to determine measured thrust function for the $T_p = 20\,°\mathrm{C}$ by interpolation between the two curves for extreme temperature. One, new, indirect method to do that is subject of this work.

2. METHOD OF INTERPOLATION

The Basic Idea of the Method

Consider the characteristics of the rocket motor for active-reactive mine 120 mm obtained by calculation by the program IBRocketMotor_M0 [4]. In figure 5 the thrust with respect to time for the three propellant temperatures – maximal value $T_p = +50\,°\mathrm{C}$, minimal value $T_p = -30\,°\mathrm{C}$, and some value between them – $T_p = 21\,°\mathrm{C}$ is shown.

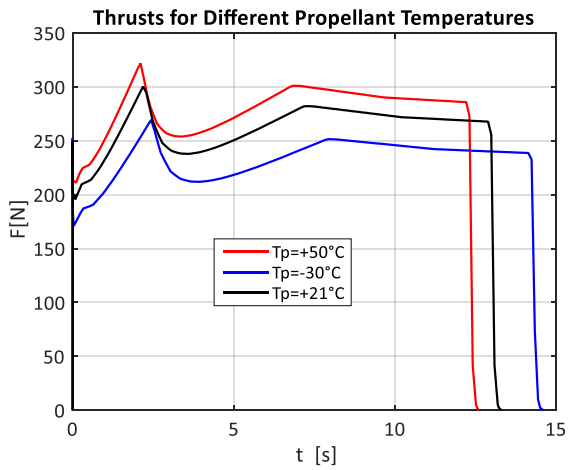


Figure 5. Calculated thrust with respect to time for the rocket motor of the AR mine.

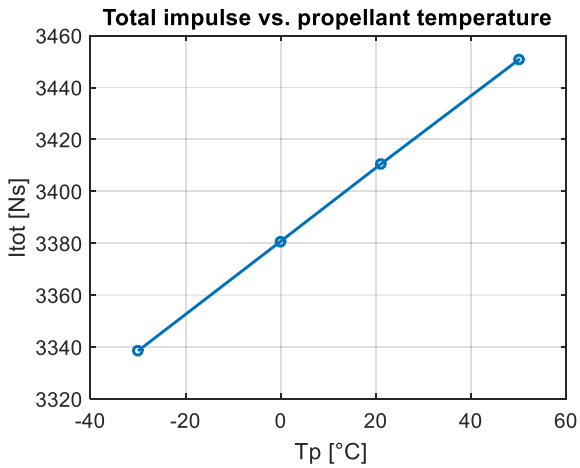


Figure 6. Total impulse with respect to propellant temperature for the rocket motor of the AR mine.

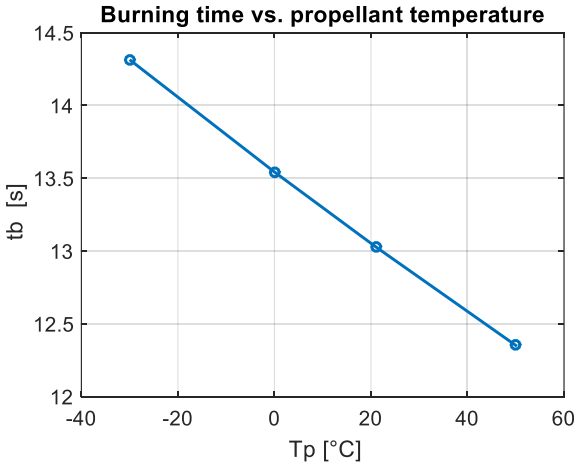


Figure 7. Burning time with respect to propellant temperature for the rocket motor of the AR mine.

For any function of thrust

$$F_i\left(T_p, t_i\right),\ i = 1,2,..,n$$

defined in n points we can calculate impulse from $t = 0$ to any point i, $i = 1,2,..,n$

$$I_t\left(T_p, t_i\right) = \mathrm{trapz}\left(t_k, F_k\right)\quad k = 1,2,..,i\ ,\ i = 1,2,..,n$$

where "trapz" denotes trapezoid method of calculation of integral. In the figure 8 the thrust with respect to impulse, and in figure 9 the burning time with respect to impulse calculated by the program IBRocketMotor_M0 [4], is shown.

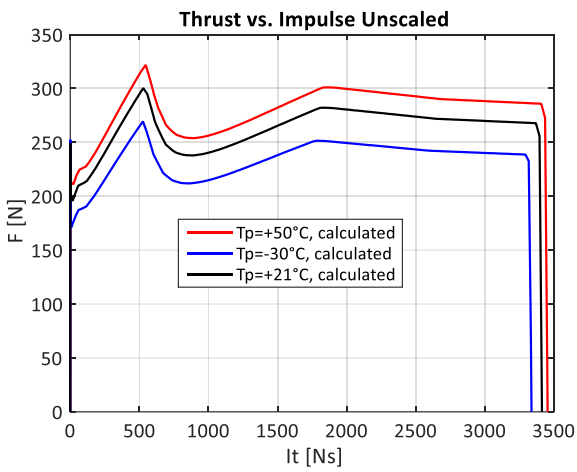


Figure 8. Thrust with respect to impulse for different propellant temperature for the rocket motor of the AR mine.

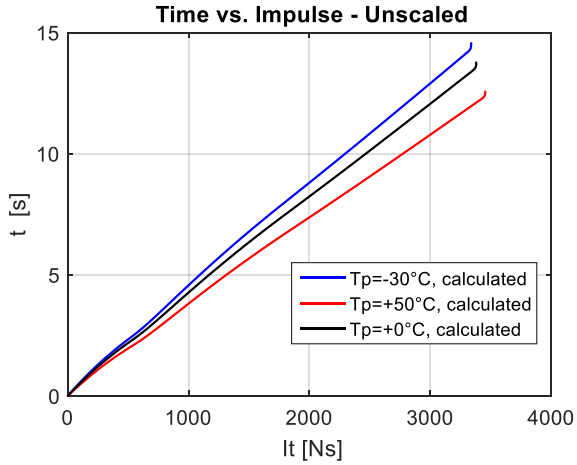


Figure 9. Burning time with respect to impulse for different propellant temperature for the rocket motor of the AR mine.

The Interpolation Procedure

If we compare the function $F(t)$ for different temperature of the propellant, shown in figure 5, with functions $F(I_t)$ in figure 8, we can conclude that thrust functions with respect to impulse almost satisfied conditions for the interpolation between different propellant temperatures. The problem is small difference in total impulse between the functions for different propellant temperatures as can be seen in figure 8. It can be solved by representing the functions $F(I_t)$ and $t(I_t)$, shown in figure 8 and 9, with respect to scaled impulse $\bar{I}_t$ for the temperature. Therefore, divide each function $I_t(T_p,t_i)$ by the total impulse for that temperature $I_{tot}(T_p)$ and obtain dimensionless scaled impulse

$$\bar{I}_{t,i}(T_p,t_i) = I_{t,i}(T_p,t_i)/I_{tot}(T_p)$$

Scaled impulse $\bar{I}_{t,i}(T_p,t_i)$ ranges between zero and unity. With scaled impulses we obtain new scaled functions $\bar{F}_i(\bar{I}_{t,i})$ and $\bar{t}_i(\bar{I}_{t,i})$ for each propellant temperature T_p. Although arrays $\bar{I}_{t,i}$ range from zero to unity for both propellant temperatures, they do not have the same discrete values. Therefore, we first have to choose the same abscissa for both boundary propellant temperatures, let say in hundred points $\bar{I}_{t,k} = 0.01, 0.02, .., 100$, for example, and then to interpolate $\bar{F}_i(\bar{I}_{t,i})$ and $\bar{t}_i(\bar{I}_{t,i})$ for both temperature for the new abscissa $\bar{I}_{t,k}$ with constant increment. It can be seen that conditions for interpolation of functions $\bar{F}_k(T_p,\bar{I}_{t,k})$ and $\bar{t}_k(T_p,\bar{I}_{t,k})$ between $\bar{F}_k(T_{p.\min},\bar{I}_{t,k})$, and $\bar{t}_k(T_{p.\max},\bar{I}_{t,k})$, $k = 1,2,..,n_m$, is now satisfied, where n_m – number of discrete points of scaled impulses. The scaled functions are shown in the figure 10 and 11 for the AR mine. Therefore, we can now interpolate thrust function and time function with respect to discrete values of scaled impulses for required propellant temperature by using any program for interpolation.

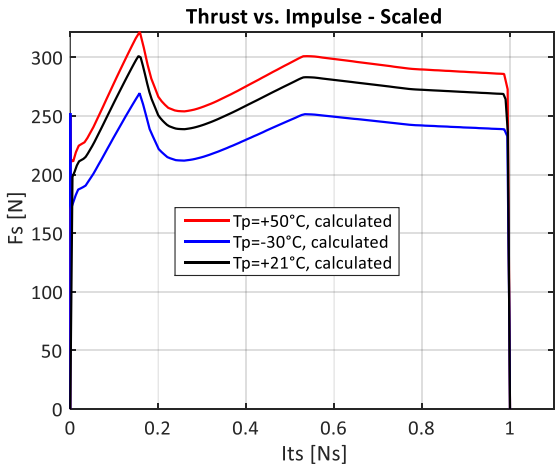


Figure 10. Thrust with respect to scaled impulse for different propellant temperature for the rocket motor of the AR mine.

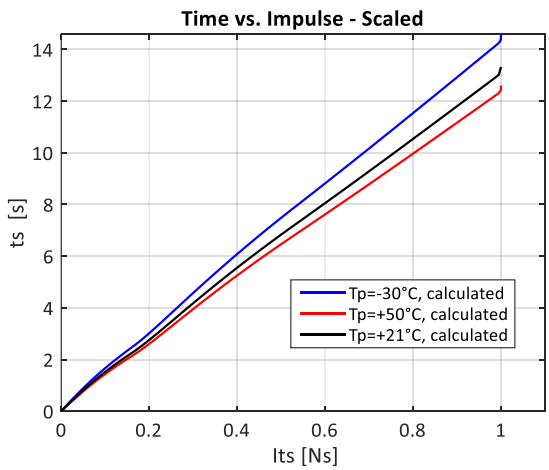


Figure 11. Burning time with respect to scaled impulse for different propellant temperature for the rocket motor of the AR mine.

After interpolation the interpolated thrust and time should be "unscaled" to actual value of total impulse for the required propellent temperature $I_{tot}(T_p)$, which is assumed known, as follows. First calculate total impulse of the interpolated thrust,

$$\bar{I}_{tot}(T_p) = \text{trapz}\left(\bar{t}_k(T_p,\bar{I}_{t,k}), \bar{F}_k(T_p,\bar{I}_{t,k})\right)$$

then calculate the ratio of the total impulses for the thrust at required temperature $I_{tot}(T_p)$ and total impulse of the interpolated scaled thrust

$$k_{T_p} = I_{tot}(T_p)/\bar{I}_{tot}(T_p),$$

and finally correct the interpolated scaled function of thrust with the coefficient of this ratio

$$F_k(T_p,t_k) = k_{T_p}\bar{F}_k(T_p,t_k), \quad k = 1,2,..,n_m$$

Coefficient of Interpolation Quality

Quality of the interpolation of thrust function can be measured by the coefficient of interpolation quality defined by

$$\bar{\sigma}_m = \sqrt{\bar{\sigma}_F^{\,2} + \bar{\sigma}_t^{\,2}}$$

where $\bar{\sigma}_F$ – Coefficient of interpolation quality of thrust

$$\bar{\sigma}_F = \frac{1}{F_{mean}} \sum_{k=1}^{n_m} \left(F_k - F_{k,m}\right)^2 \Big/ \left(n_m - 1\right)$$

and $\bar{\sigma}_t$ – Coefficient of interpolation quality of time

$$\bar{\sigma}_t = \frac{1}{t_b} \sum_{k=1}^{n_m} \left(t_k - t_{k,m}\right)^2 \Big/ \left(n_m - 1\right)$$

In the expression above quantities F_k and t_k are interpolated thrust and time, and $F_{k,m}$ and $t_{k,m}$ are measured thrust and time, n_m – number od data, and F_{mean} is mean value of thrust. Note that total impulse of the interpolated thrust is equal to the total impulse of the measured thrust for the cases where we compare interpolated thrust function with the known measured one.

Consideration of the Calculated Example

Based on explained procedure the program ThrustInterpolation for interpolation of thrust function is made [5]. Program does linear interpolation, but higher order can be implemented. Calculated values of thrust functions for the extreme temperatures we can consider as experimentally obtained values (figure 12).

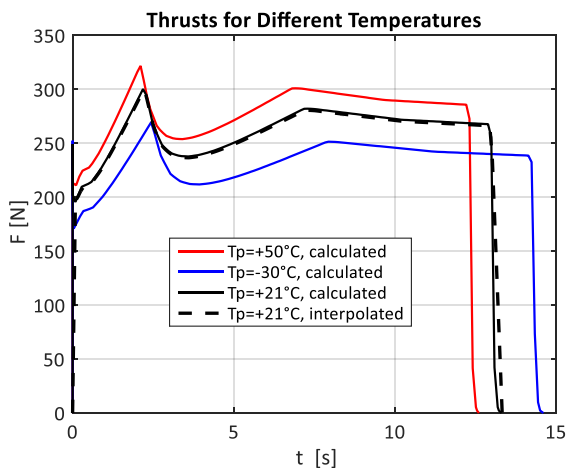


Figure 12. Interpolated thrust in comparison of the calculated for the rocket motor of the AR mine.

Based on these values and explained procedure we interpolated the thrust function for propellant temperature $T_p = 21\,°\mathrm{C}$. In the figure 12 interpolated thrust is represented by dushed line. On the same diagram the calculated value for propellant temperature $T_p = 21\,°\mathrm{C}$ is shown for the comparison. We can see excellent matching between calculated and interpolated curves. Calculated coefficient of interpolation quality is $\bar{\sigma}_m = 3.0\,\%$.

3. TESTING THE METHOD ON REAL EXAMPLES

AT Rocket 90 mm M79 OSA

In the introduction of this report, we stated basic characteristics of the anti-tank shoulder launch rocket 90 mm M79 OSA.

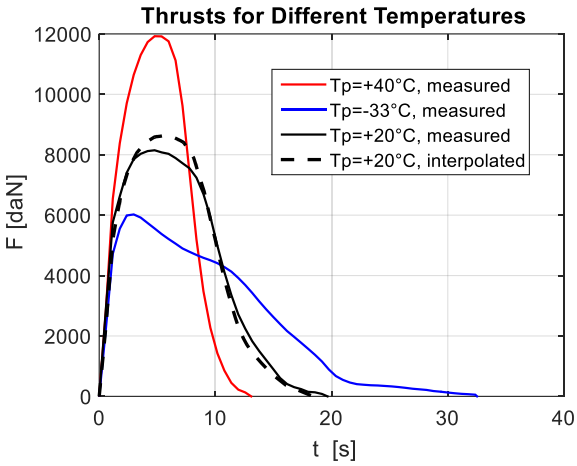


Figure 13. Interpolated thrust in comparison of the calculated for the AT rocket OSA.

Now let us interpolate thrust function for the propellant temperature $T_p = 20\,°\mathrm{C}$ for which measurement is available. By using the program ThrustInterpolation the interpolated function is calculated and shown in figure 13 in parallel with measured values. Taking in mind highly nonlinear dependance of the total impulse with respect to propellant temperature and large erosion in burning the propellant we can conclude that matching between calculated and interpolated curves are very good. Coefficient of interpolation quality is $\bar{\sigma}_m = 4.8\,\%$.

ART Rocket R-2

For the ART rocket R-2 we have measurement of the thrust

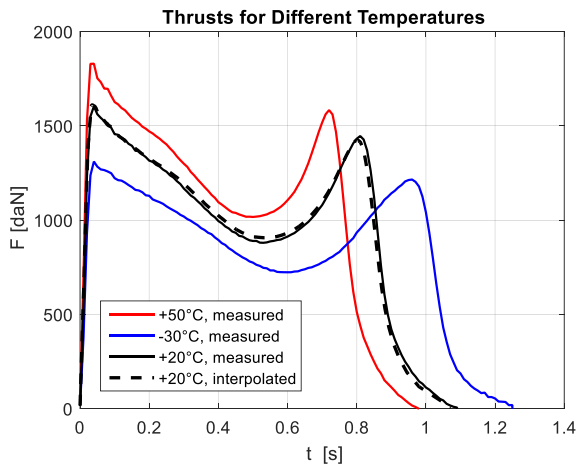


Figure 14. Interpolated thrust in comparison with the measured for the ART rocket.

for the three temperatures, (figure 14), maximum minimum and one at $T_p = 20°C$. Interpolated value for that temperature is also shown in the same figure.

We can see excellent matching between calculated and interpolated curve. Coefficient of interpolation quality is $\bar{\sigma}_m = 1.8\%$.

Guided Anti-tank Weapon BUMBAR

For the Guided Anti-tank Weapon BUMBAR available maximal value of the propellant temperature is $T_p = +50°C$, minimal value $T_p = -30°C$, and two values between them – $T_p = +35°C$ and $T_p = +20°C$ are also available, [6]. The thrusts in time are shown in figure 15. and 16. Therefore, we can test the algorithm in the four combinations.

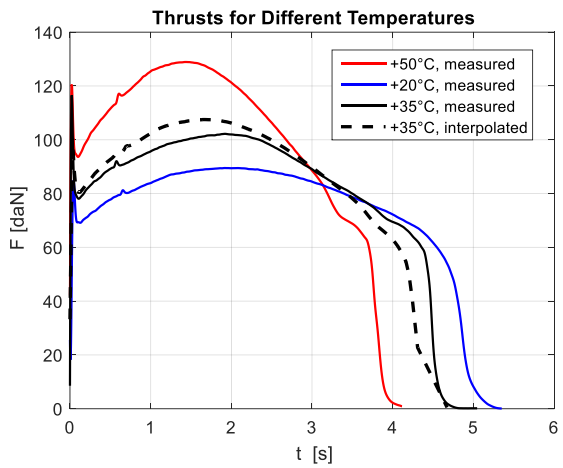


Figure 15. Interpolated thrust in comparison of the measured for the GAW BUMBAR, T_p =+35°C.

In figure 15 and 16 two results of interpolation are shown in parallel with corresponding measurements. Comparison of the interpolated and the measurement curves shows relatively good matching taking in mind relatively large erosive burning in rocket motor, [6]. Coefficient of interpolation quality for the interpolation of thrust for $T_p = +35°C$ is $\bar{\sigma}_m = 5.6\%$ and for $T_p = +20°C$ is $\bar{\sigma}_m = 4.2\%$.

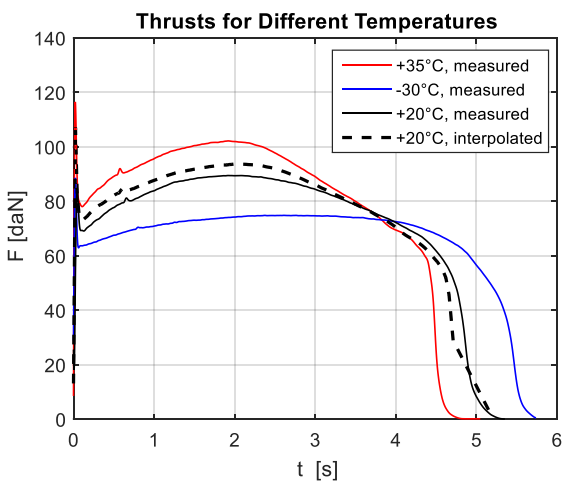


Figure 16. Interpolated thrust in comparison with the measured for the GAW BUMBAR, T_p =+20°C.

Sustainer Rocket Motor of the SAM R-3

For the sustainer rocket motor of the guided SAM R-3 maximal value of propellant temperature is $T_p = +50°C$, minimal value $T_p = -30°C$, and two values between them: $T_p = +18°C$ and $T_p = +0°C$ is shown in figure 17.

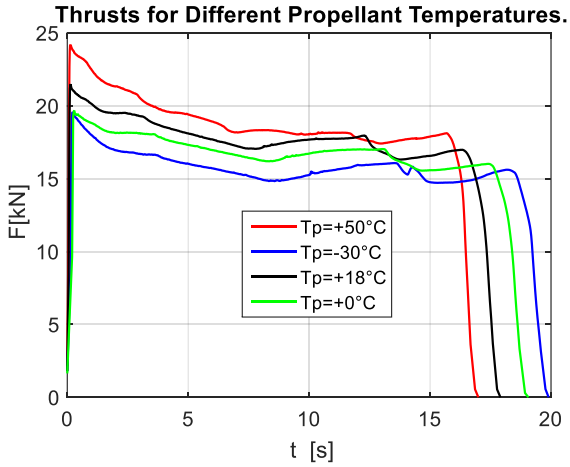


Figure 17. Measured thrusts of the rocket motor SAM R-3 for the four different propellant temperatures.

Let us interpolate thrust function for the powder temperature of $T_p = +18°C$ on two ways – first between curves for $T_p = +50°C$ and for $T_p = -30°C$ and the second between curves for $T_p = +50°C$ and for $T_p = 0°C$.

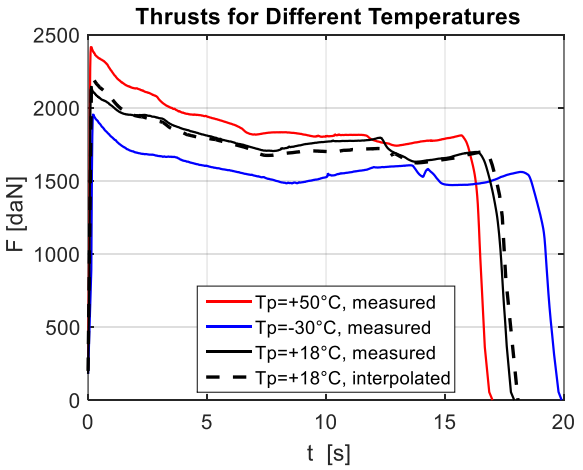


Figure 18. Interpolated thrust for T_p =+18°C between measurement T_p =+50°C and T_p =-30°C in comparison with the measured value for the rocket motor SAM R-3.

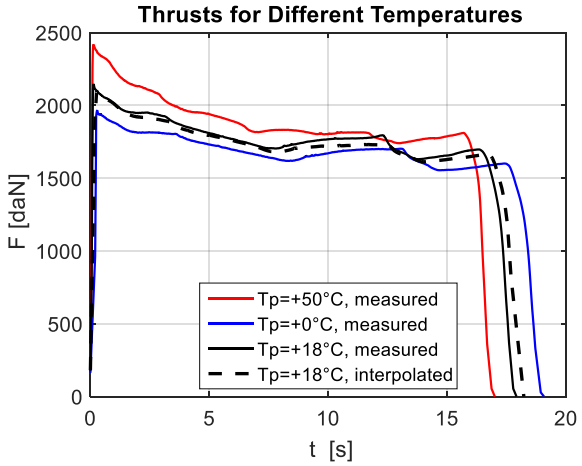


Figure 19. Interpolated thrust for T_p =+18°C between measurement T_p =+50°C and T_p =0°C in comparison with the measured values for the rocket motor SAM R-3.

The interpolated thrust function for the powder temperature $T_p = +18\,°\mathrm{C}$ between curves for $T_p = +50\,°\mathrm{C}$ and for $T_p = -33\,°\mathrm{C}$ is shown in figure 18, and in figure 19. between curves for $T_p = +50\,°\mathrm{C}$ and for $T_p = 0\,°\mathrm{C}$. Coefficient of interpolation quality is $\bar{\sigma}_m = 2.3\%$ for powder temperature $T_p = +18\,°\mathrm{C}$ and $\bar{\sigma}_m = 2.7\%$ powder temperature $T_p = 0\,°\mathrm{C}$.

Again, we can see the matching between interpolated and measured curves are fairly good for both approaches.

Guided Anti-tank Weapon MALYUTKA

For the Guided Anti-tank Weapon MALYUTKA, for the start motor, data are available for the maximal value of the propellant temperature $T_p = +50\,°\mathrm{C}$, minimal value $T_p = -30\,°\mathrm{C}$, and for the value between them – $T_p = +20\,°\mathrm{C}$ are also available, [6]. The thrusts in time are shown in figure 20.

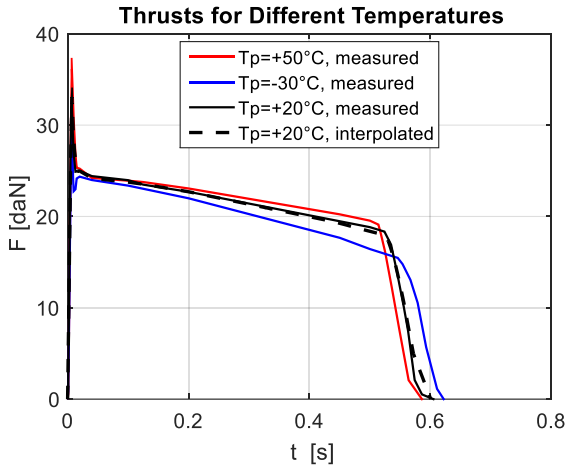


Figure 20. Interpolated thrust in comparison of the measured for the GAW MALYUTKA, T_p =+20°C.

We can see excellent matching between measured and interpolated thrust for $T_p = +20\,°\mathrm{C}$. Coefficient of interpolation quality is $\bar{\sigma}_m = 2.4\%$.

Artillery Rocket 122 mm

For the artillery rocket 122 mm data are available for the maximal value of the propellant temperature $T_p = +50\,°\mathrm{C}$, minimal value $T_p = -40\,°\mathrm{C}$, and for the value between them – $T_p = +20\,°\mathrm{C}$ are also available. The thrusts in time are shown in figure 21.

At minus temperature the thrust is almost constant. At plus maximum propellant temperature the thrust is almost constant after 0.25 s, but the maximum value of the thrust at the beginning of operation of motor is about fifty percent greater than that constant value. This is most likely due to high erosive burning at the beginning of the operation of motor. Even though we can see reasonable good interpolated function for the propellant temperature $T_p = +20\,°\mathrm{C}$ Coefficient of interpolation quality is $\bar{\sigma}_m = 6.4\%$.

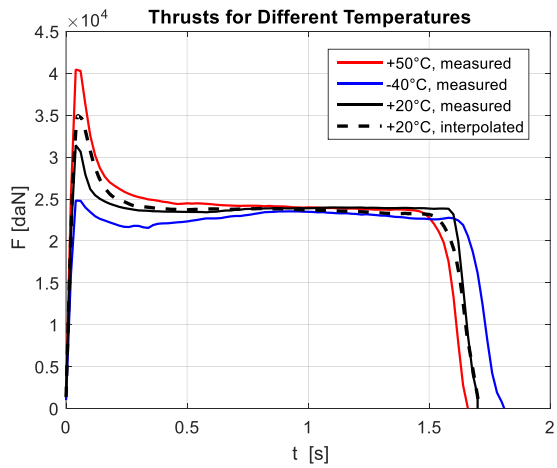


Figure 21. Interpolated thrust in comparison of the measured for the ART rocket, T_p =+20°C.

4. CONCLUSION

Proposed new numerical method to interpolate the thrust curve for a given propellant temperature, based on the known, measured thrust functions for one higher and one lower temperature, gives very good results for engineering application. The method is proved based, first on calculation, and then based on several data obtained by measurement of the rocket motor characteristics for different types of rockets and propellant temperature.

Method gives better results for the cases where burning of propellant is without erosive burning, but it can be used even at high level of erosive burning.

References

[1] Sutton,H., Seifert,G.P.: Rocket Propulsion Elements, Wiley, 1963.

[2] Јовановић,В.: *Прича о ОСИ*, Академска мисао, Београд, 2026.

[3] Ćurčin,M.: Program ThrustAnalyzerMulti, Beograd, 2015 (private property).

[4] Jovanović,V, Ćurčin,M.: Program IBRocketMotor, Beograd, 2012 (private property).

[5] Ćurčin,M.: Program ThrustInterpolation, Beograd, 2025 (private property).

[6] Stojković,S., Gligorijević,N.: *Effect of erosive burning in rocket motor on the ballistics of rocket flight*, OTEH 2024, Tara, Serbia, 2024

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ANALYSIS OF RECOIL RESISTANCE FORCE BEHAVIOR AS A FUNCTION OF VARIATIONS IN INFLUENTIAL PARAMETERS

MILAN SIMONOVIĆ
Military Technical Institute, Belgrade, milan.simonovic@mod.gov.rs

MARKO SIMONOVIĆ
Military Technical Institute, Belgrade, marko.simonovic@mod.gov.rs

NIKOLA RADOVANOVIĆ
Military Technical Institute, Belgrade, nikola.radovanovic@mod.gov.rs

ANDRIJA VELJOVIĆ
Military Technical Institute, Belgrade, andrija.veljovic@mod.gov.rs

ANA SIMONOVIĆ
Military Technical Institute, Belgrade, ana.gacic@mod.gov.rs

SANJA MILINKOVIĆ
Military Technical Institute, Belgrade, sanja.milinkovic@mod.gov.rs

Abstract: *In this paper, the dynamic behavior of a hydraulic recoil mechanism during the firing process is analyzed by varying key parameters that influence the maximum recoil velocity and recoil length, the maximum recoil braking force and the total recoil resistance force. The behavior of the recoil resistance force was analyzed at an elevation angle of 45°, considering variations in the size of the flow control orifices. Additionally, the values of the additional (supplementary) mass and the total recoiling mass were varied. Following the parameter variations, the obtained results were analyzed individually for each case, as well as comparatively across all cases, in order to derermine the optimal combination of parameters that does not adversely affect the operation of the hydraulic recoil mechanism and consequently the stability of the artillery system.*

Keywords: *hydraulic recoil mechanism, barrel recoil, mortar, analysis.*

1. INTRODUCTION

Modern artillery systems require high levels of firing accuracy, reliability, and operational effectiveness. One of the most important stages in the design of artillery systems is the determination of the motion characteristics of their recoiling mass during projectile launch. Upon firing, the projectile and the powder charge acquire a certain amount of momentum, which is transferred to the launching system, i.e. to the weapon. If there were no elastic connection between the barrel and the carriage, the entire weapon would, at the instant the projectile leaves the barrel, receive the same amount of momentum, resulting in an impulsive load of very short duration and an extremely high mechanical force. This would cause the entire weapon to recoil, requiring the weapon to be respositioned and re-aimed at the selected target.

To prevent these adverse effects, the recoil impulse is transmitted to the recoiling parts of the weapon through the recoil mechanism. The action of the recoil mechanism slows down the recoil motion and prolongs the duration of the load acting on the carriage, while also reducing the mechanical force to an acceptable level (from the standpoint of weapon stability and structural integrity). The dynamics of the recoiling parts must be such that, during the projectile's travel through the barrel, they generate the smallest possible oscillations that could compromise firing precision and accuracy [1].

For large-caliber artillery systems, such as a 203 mm weapon with a smooth barrel, the recoil length is one of the key parameters directly affecting the design of the recoil mechanism, the loading of the carriage, and the overall performance of the system. The recoil length is influenced by a number of related parameters, including projectile characteristics, the powder charge, the mass of the moving parts, as well as the parameters of the hydraulic and pneumatic components of the recoil system.

In engineering practice, the design of artillery systems over the past decades has relied on conventional methods for determining the motion of recoiling parts during both recoil and counter-recoil. These methods are based on the approximate integration of the differential equations governing the motion of the recoiling mass, with the introduction of appropriate simplifying assumptions and hypotheses that enable the integration.

Corresponding author: M. Simonović;
E-mail: milan.simonovic@mod.gov.rs
DOI: 10.5937/OTEH26026S

Advances in information technology and computer engineering have created the conditions for the numerical integration of differential equations and have enabled more sophisticated analyses of the behavior of artillery systems during projectile launch. Such analyses also make it possible to investigate the dynamic behavior of the weapon during firing and to perform a more detailed assessment of its structural integrity under dynamic and impulsive loading conditions [1].

For artillery systems equipped with an elastic carriage, stability during the firing process must be ensured. To achieve the required stability, an appropriate variation of the braking force during recoil and counter-recoil is necessary. Furthermore, it is required to establish the relationship between the velocity and displacement of the recoiling parts and the forces P_{kn} and R suring recoil, as well as the force R_v during counter-recoil, as functions of time or the displacement of the recoiling parts. Since no analytical lav describing the variation of $P_{kn}(t)$ exists, the kinematic quantities of the recoiling parts (V, X, y, t) cannot be determined by integrating the differential equation of motion of the recoiling parts [2].

Therefore, the analysis of the motion of the recoiling parts compromises the following three phases:

- the free recoil phase, i.e. motion under the action of the force P_{kn} without any braking;
- the braked recoil phase, i.e. motion under the action of the braking force R and
- the counter-recoil phase, i.e. motion under the action of the force R_v.

The motion of the recoiling parts during free recoil begins together with the projectile motion inside the barrel. To facilitate the calculations, the recoil process is divided into three periods:

- the first period (projectile travel through the barrel);
- the second period (outflow of powder gases) and
- the third period (inertial motion of the recoiling parts, P_{kn}=0) [2].

The kinematic quantities characterizing free recoil are the velocity W and the displacement L.

Braked recoil represents the actual motion of the recoiling mass during the firing process under the action of the forces acting on the recoiling mass (the force P_{kn} and the resultant recoil resistance force R). The motion of the recoiling parts during braked recoil is analyzed over the same periods as those considered for free recoil [2].

The law governing the variation of the resultant recoil resistance force R(t) is specified in advance, depending on the type of artillery system, and must satisfy the following requirements:

- the maximum value of the recoil resistance force must not exceed its maximum allowable value determined by the stability requirements of the weapon during recoil;
- the braking of the recoiling parts must be uniform, and their kinetic energy must be completely dissipated over the specified recoil length within the specified time interval;
- the loading of the carriage during firing must be kept as low as possible [2].

2. INPUT INTERNAL BALLISTIC PAREMETERS

The internal ballistic analysis is used to determine the maximum powder gas pressure (P_{max}), the propellant gas pressure at the muzzle (P_u), and the projectile muzzle velocity (v_0) [3]. The input data for the internal ballistic analysis are presented in Table 1. For the purposes of the analysis, a single type of powder was used.

Table 1. Input data for the internal ballistic analysis

No.	Characteristic parameter	Value
1.	Barrel cross-sectional area, S [m^2]	0.032360
2.	Total forward travel of the projectile in the barrel, X_u [m]	3.34
3.	Initial volume of the powder chamber, W_0 [m^3]	0.02150
4.	Cross-sectional area of the clearance between the projectile body and the barrel wall, S_z [m^2]	0.000001
5.	Clearance shape coefficient, ξ	0.61
6.	Projectile mass, m [kg]	100
7.	Projectile cross-sectional area, S_{pr} [m^2]	0.032360
8.	Specific work of the powder gases of the supplementary powder charge, f_b [J/kg]	1080000
9.	Covolume of the powder gases of the supplementary powder charge, α	0.00077
10.	Powder density, ρ [kg/m^3]	1600
11.	Burning rate coefficient of the supplementary powder charge, u_{z0} [m/sPa]	$6.158 \cdot 10^{-8}$
12.	Ratio of specific heats, κ	1.2
13.	Half-thickness of the powder grain, r_0 [m]	0.00026
14.	Shape coefficient, κ_1	1.2026
15.	Shape coefficient, λ_1	-0.16846
16.	Specific work of the powder gases of the main powder charge, f_{b0} [J/kg]	1080000
17.	Covolume of the powder gases of the main powder charge, α_0	0.00077
18.	Mass of the one bag of the supplementary powder charge, m_{dbp} [kg]	0.7
19.	Mass of the main powder charge, m_{obp} [kg]	0.078

The muzzle velocity and the maximum pressure depend on the mass of the powder charge, which was varied by adding one to six bags of the supplementary powder charge. Table 2 presents the values of the muzzle velocity and the maximum powder gas pressure as functions of the powder charge mass and the projectile mass, while their variations are illustrated in Figures 1 and 2.

Table 2. Values of the muzzle velocity and maximum powder gas presure

Powder charge mass	1	2	3	4	5	6
m_{bp} [kg]	0.778	1.478	2.178	2.878	3.578	4.278
V_0 [m/s] 75 kg	176.23	248.06	303.54	352.21	394.80	434.29
P_m [bar] 75 kg	219.59	516.21	836.54	1176	1534.26	1912.05
V_0 [m/s] 100 kg	154.49	215.55	263.32	304.96	342.09	376.58
P_m [bar] 100 kg	244.91	547.78	870.51	1213.4	1573.66	1953.24
V_0 [m/s] 125 kg	138.11	192.32	235.91	272.86	306.45	336.87
P_m [bar] 125 kg	261.29	567.54	893.31	1236.34	1597.73	1978.33

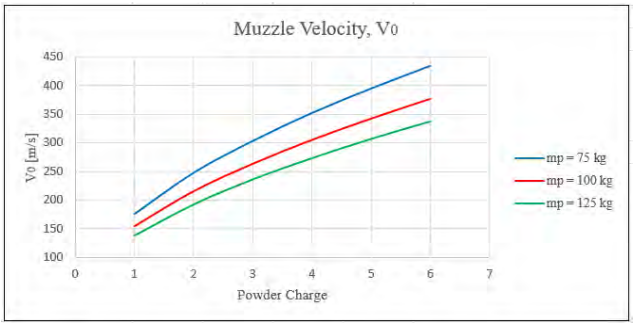


Figure 1. Variation of the projectile muzzle velocity as a function of the amount of the supplementary powder charge

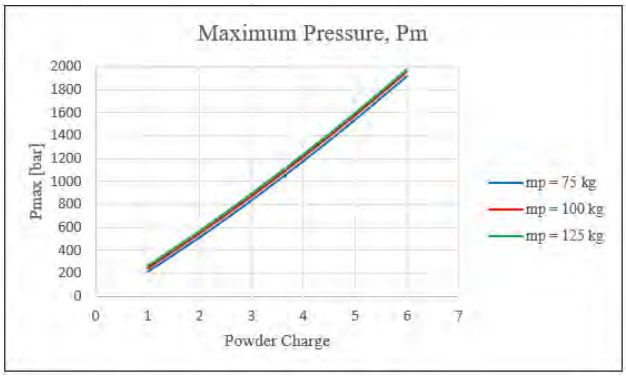


Figure 2. Variation of the maximum propellant gas pressure as a function of the amount of the supplementary powder charge

3. CALCULATION OF THE RECOIL RESISTANCE FORCE

To investigate the possibility of achieving the desired law governing the variation of the recoil resistance force, it is necessary to analyze its components. These components are: the recuperator force (F_p), the recoil brake force (F_{kt}), the friction force acting on the cradle guideways and the seals (F_{tr}) and the projection of the weight of the recoiling parts onto the direction of recoil ($M_0 g \sin\varphi$). The general equation for determining the recoil resistance force at an elevation angle φ is given by [1]:

$$R = F_p + F_{kt} + F_{tr} - M_0 g \sin\varphi \tag{1}$$

The recoil resistance force acts along the line passing through the center of mass of the recoiling parts (parallel to the bore axis) and is independent of the arrangement of the recoil mechanism components on the weapon. The approximate value of the friction force acting on the cradle guideways and in the seals of the recoil mechanism is given by the following equation [1]:

$$F_{tr} = M_0 g (f \cos\varphi + \nu) \tag{2}$$

where:

f - coefficient of friction i ν - proportionality coefficient for the seal fricition force.

From the diagram illustrating the variation of the forces acting on the recoiling parts during recoil (Figure 3), it can be observed that the desired braking law can be achieved only by an appropriate selection of the hydraulic resistance force in the recoil brake. For engineering design calculations, the friction forces may be considered constant throughout the entire recoil stroke with sufficient accuracy. The recuperator force is determined based on the requirement that the recoiling mass returns to its forward position after the completion of recoil and remains in that position at all elevation angles. Therefore, the only parameter that can be adjusted is the law governing the variation of the force in the hydraulic recoil brake.

Consequently, one of the fundamental requirements in the design of the recoil brake is to achieve the prescribed variation of its resistance force throughout the recoil stroke [1].

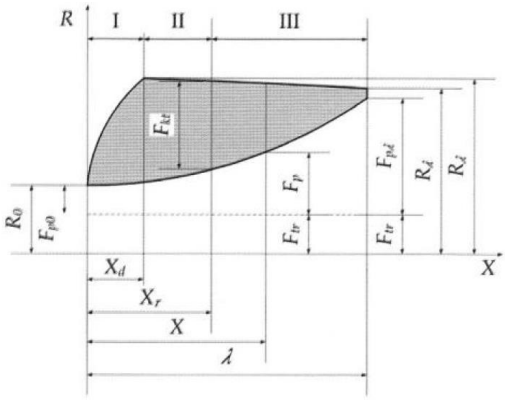


Figure 3. Components of the recoil resistance force [1]

Furthermore, it is well established that, for artillery systems with given energy characteristics and a prescribed recoil length, the loading of the carriage is minimized when the braking of the recoiling parts is achieved by a constant braking force throughout the entire recoil stroke.

It should be borne in mind that all design calculations are performed for specified and known operating conditions of the artillery system. If any of the operating conditions change, the system behavior and structural integrity must be re-evaluated under the new conditions [1].

The initial value of the recoil resistance force is determined by the following equation:

$$R_0 = F_{p0} + F_{tr} - M_0 g sin\varphi \quad (3)$$

4. ANALYSIS OF THE HYDRAULIC RECOIL BRAKE PERFOMANCE UNDER VARIATIONS OF INFLUENTIAL PARAMETERS

To provide a clearer evaluation of the behavior of the hydraulic recoil brake during the firing process under different operating conditions, the following parameters were varied: the powder charge (by adding from one to six bags of powder, each bag containing 700 g of the same type of powder), the projectile mass, m_p (75, 100 and 125 kg), the recoiling mass, m_t (the existing recoiling mass increased by 100, 200 and 300 kg, as well as the original mass without any increase), and the flow area of the ports in the hydraulic recoil brake, which is determined by the positions of the long and short grooves in the recoil brake counter-piston. The positions of the long and short grooves are adjusted by rotating the counter-piston, thereby regulating their opening percentages. The analysis was carried out for a 203 mm artillery system with a smooth barrel. An elevation angle of 45° was selected. Table 3 presents the combinations of the opening of the long and short grooves used in the analysis.

Table 3. Combinations of the opening percentages of the long and short grooves

Long	100%	100%	100%	100%	Long
Short	0%	25%	50%	75%	100%
Long	0%	25%	50%	75%	Short 100%
Short	100%	100%	100%	100%	

Figures 4-7 present the maximum recoil length, X_{max}, the maximum recoil velocity, V_{max}, the maximum recoil resistance force, R_{max} and the maximum hydraulic recoil brake force, F_{kmax}, for all previously described paremeter combinations.

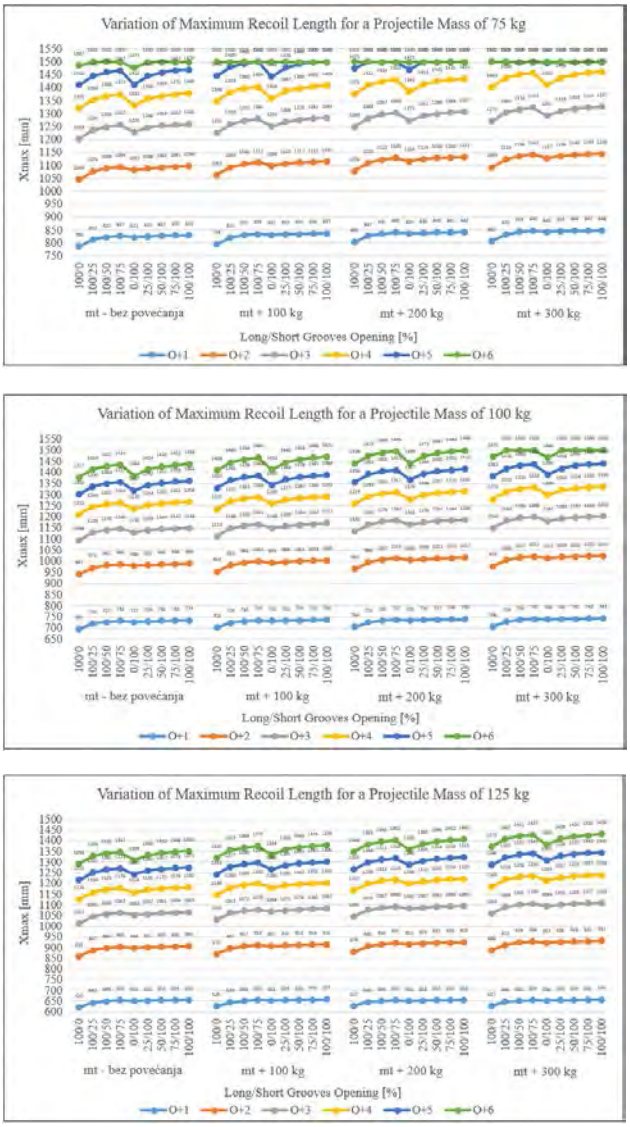


Figure 4. Maximum recoil length

The analysis of the diagram showing the maximum recoil length (X_{max}) indicates that, in all cases considered, the minimum recoil length is achieved with the combination of a fully open long groove and a fully closed short groove. This trend remains unchanged independent of the projectile mass, the recoiling mass, and the powder charge used. As the opening percentage of the short groove increases, the maximum recoil length gradually increases.

When the effects of the recoiling mass and the projectile mass are considered, it can be observed that a 75 kg projectile exceeds the acceptable recoil length limit for all

supplementary powder charges. Without an increase in the recoiling mass, firing is permissible up to supplementary powder charge No. 5, while for configurations with an increased recoiling mass, the upper permissible limit is supplementary powder charge No. 4.

For a 100 kg projectile, acceptable recoil length values for all six supplementary powder charges are achieved when the recoiling mass is unchanged or increased by 100 kg or 200 kg. However, for safety reasons and to ensure the protection of the weapon system, firing should be carried out using the configuration with a fully open long groove and a fully closed short groove.

An increase in the recoiling mass by 300 kg limits the use of all supplementary powder charges, as the highest supplementary charge must not be used because it results in the maximum permissible recoil length being reached.

With a 125 kg projectile, the most favorable results are obtained under all operating conditions considered. Nevertheless, even in this case, firing with the long grooves fully open and the short grooves fully closed in the recoil brake counter-piston is recommanded.

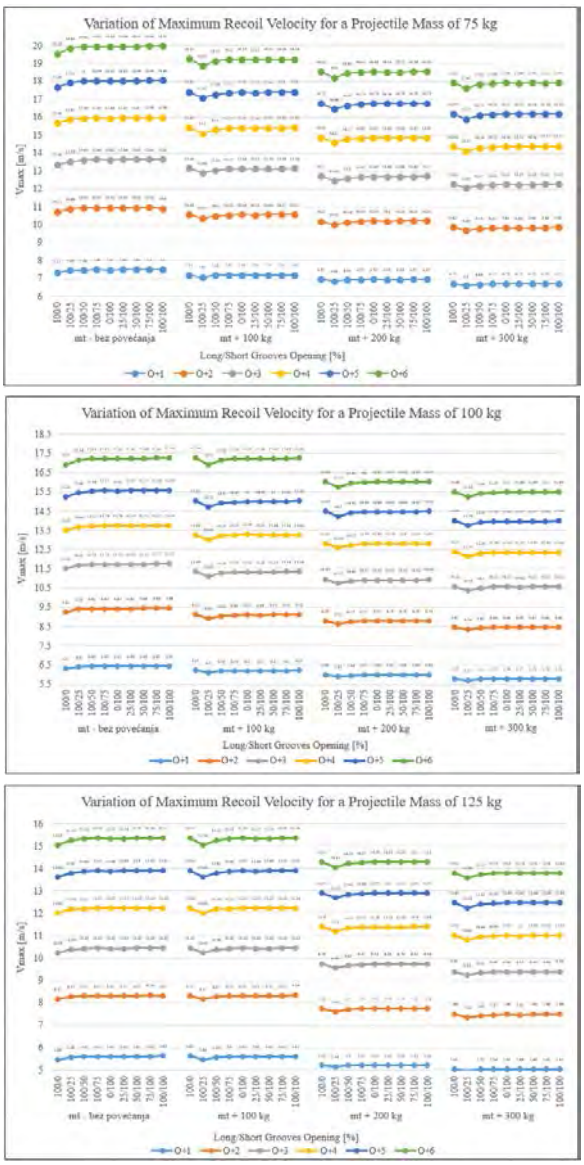


Figure 5. Maximum recoil velocity

In all cases considered, the largest recoiling mass results in the lowest maximum recoil velocity (V_{max}). An increase in the projectile mass leads to a reduction in the maximum recoil velocity, while an increase in the barrel mass further decreases the maximum recoil velocity. The highest recoil velocities are obtained for a 75 kg projectile and the lowest barrel mass, whereas the lowest recoil velocities are achieved for a 125 kg projectile and the highest barrel mass

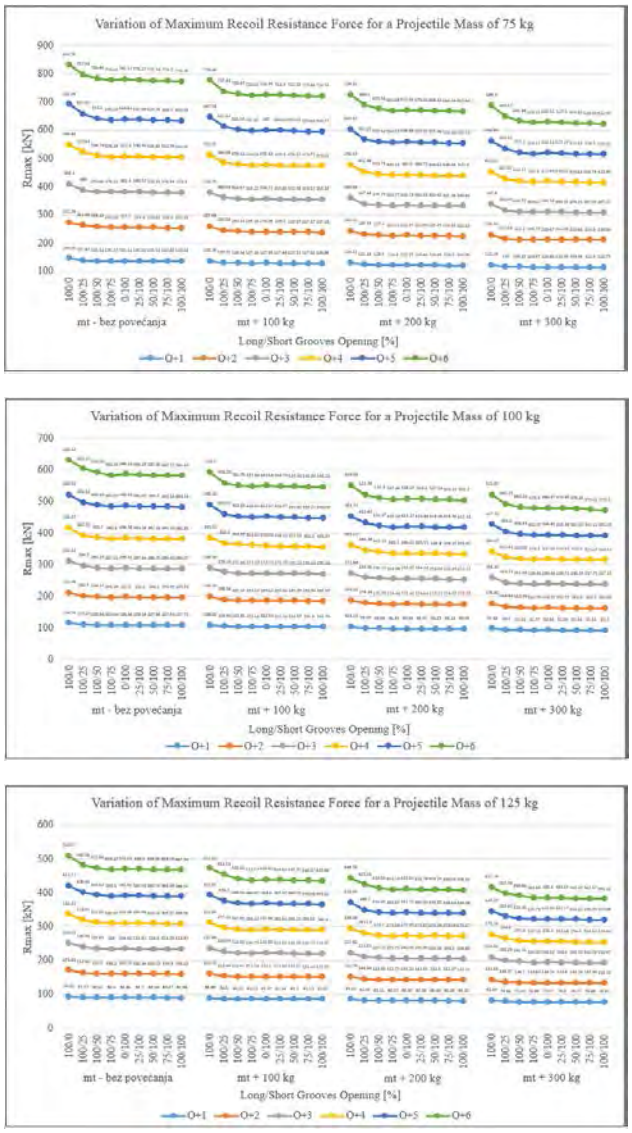


Figure 6. Maximum recoil resistance force

As the recoiling mass and the projectile mass increase, the maximum recoil resistance force (R_{max}) decreases. The maximum recoil resistance force is achieved with the combination of a fully open long groove and a fully closed short groove in the recoil brake counter-piston, whereas the minimum value is obtained when both grooves are fully open. It can also be observed that increasing the amount of the supplementary powder charge results in a significant increase in the maximum recoil resistance force.

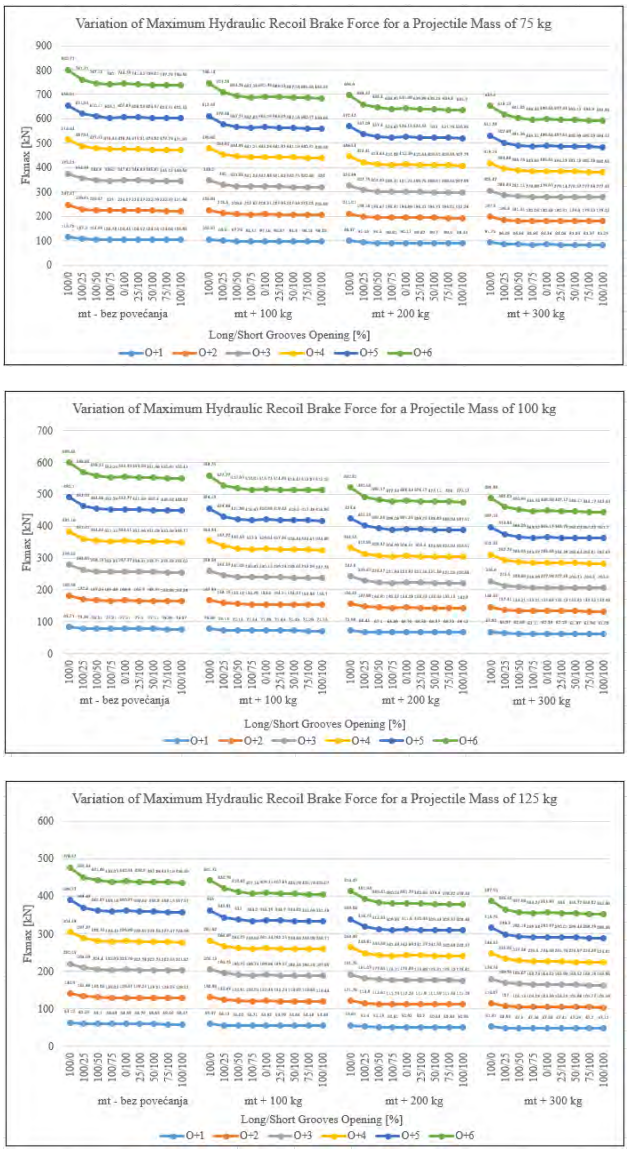


Figure 7. Maximum hydraulic recoil brake force

The maximum hydraulic recoil brake force (F_{kmax}) decreases with increasing projectile mass as well as with increasing recoiling mass. The powder charge has a significant influence on the maximum hydraulic recoil brake force, which increases significantly with an increase in the powder charge mass.

5. OPTIMIZATION OF THE INPUT PARAMETERS

Based on the conducted analysis of the effects of the projectile mass, the recoiling mass, and the grooves opening configuration of the hydraulic recoil brake counter-piston on the fundamental parameters governing the performance of the recoil brake (maximum recoil length X_{max}, maximum recoil velocity V_{max}, maximum recoil resistance force R_{max} and maximum hydraulic recoil brake force F_{kmax}), the system was optimized with the objective of enabling the use of all six powder charges without causing damage to the 203 mm artillery system.

The analysis of the obtained results showed that the counter-piston channel configuration with the long grooves fully open and the short grooves fully closed provides the most favorable operating characteristics of the system. This configuration yields the minimum values of the maximum recoil length and maximum recoil velocity while also producing the maximum values of the recoil resistance force and the hydraulic recoil brake force, indicating the most efficient deceleration of the recoiling parts during the recoil process. Therefore, this configuration was adopted as the optimal design solution.

During the optimization process, an additional design constraint was adopted, according to which the maximum recoil length must not exceed 1450 mm, although the structurally permissible limit is 1500 mm. The provides an additional 50 mm safety reserve, thereby enchancing the reliability of the system under operating conditions.

The results showed that the 75 kg projectile does not satisfy the established requirement for the highest powder charge (O+6), since the maximum recoil stroke exceeds the adopted limit. In contrast, the 100 kg and 125 kg projectiles satisfy the specified requirement, with the 125 kg projectile exhibiting the lowest values of all analyzed parameters.

It is particularly important that the 125 kg projectile, in combination with the configuration featuring fully open long grooves and fully closed short gloves, and without any additional increase in the recoiling mass, achieves a maximum recoil stroke of 1292 mm with the O+6 powder charge, providing a 158 mm safety reserve with respect to the adopted limit of 1450 mm.

At the same time, this configuration yields the lowest values of both the maximum recoil resistance force and the maximum hydraulic recoil brake force compared with all other analyzed configurations.

Based on the overall analysis, the optimal system configuration was identified as the combination of a 125 kg projectile, a counter-piston configuration with the long grooves fully open and the short grooves fully closed, without any increase in the recoiling mass. This configuration enables the use of all six additional powder charges while satisfying the established design constraints and providing the most favorable operating characteristics of the recoil system.

A particularly important advantage of the proposed optimal solution is that it is achieved without increasing the recoiling mass, thereby avoiding any additional structural loading of the system. Table 4 presents the values of the hydraulic recoil brake performance indicators for the selected optimal configuration.

Table 4. Performance indicators of the hydraulic recoil brake for the selected optimal configuration

	O+1	O+2	O+3	O+4	O+5	O+6
X_{max} [mm]	620	859	1013	1126	1216	1292
V_{max} [m/s]	5.48	8.17	10.23	12.02	13.62	15.04
R_{max} [kN]	94.01	171.45	252.03	336.41	422.21	510.07
F_{kmax} [kN]	63.12	140.9	220.15	304.49	389.77	476.52

6. CONCLUSION

This paper analyzes the behavior of the hydraulic recoil system of a 203 mm artillery system, with a smooth barrel with respect to the main parameters of the recoil process. Following the presentation of the theoretical background and the internal ballistic parameters used in the calculations, a numerical analysis was performed to investigate the influence of different recoil brake flow-grooves configurations, projectile mass, and additional recoiling mass on the main performance characteristics of the system, including the maximum recoil length, maximum recoil velocity, maximum recoil resistance force and maximum hydraulic recoil brake force.

The obtained results indicate that increasing the projectile mass leads to a reduction in all previously analyzed parameters describing the performance of the hydraulic recoil brake. It was also found that the selected combination of long and short grooves openings provides the most favorable operating characteristics of the system by achieving the lowest values of the maximum recoil velocity and maximum recoil length, while while also ensuring the highest braking efficiency of the recoiling mass.

Based on the optimization results, the optimal configuration consists of a 125 kg projectile, with the long grooves fully open and the short grooves fully closed without increasing the recoiling mass. This configuration enables the use of all six analyzed powder charges while satisfying the maximum allowable recoil length of 1450 mm and providing a sufficient design safety reserve with respect to the permissible travel limit of the recoiling parts under the most critical operating conditions.

The results demonstrate that an appropriate selection of the flow-grooves geometry and the main mass-related parameters can significantly improve the performance of the recoil system without increasing the mass of the recoiling components. Consequently, reliable operation of the artillery system is ensured for all analyzed powder charges, while reducing structural loading and maintaining system stability throughout the firing process.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03-34/2026-03/200325).

References

[1] KALEZIĆ, M.: *Projektovanje artiljerijskih sistema*, Knjiga jedan, ISBN 978-86-912671-0-0 Beograd, 2010.

[2] RISTIĆ, Z., JAKOVLJEVIĆ, M.: *Osnovi naoružanja-konstrukcija artiljerijskog oruđa*, Knjiga, Vojna akademija, Beograd, 2001.

[3] CVETKOVIĆ, M.: *Unutrašnja balistika*, Knjiga, Sektor za školstvo, obuku, naučnu i istraživačku delatnost, Beograd, 1998.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DYNAMIC ANALYSIS OF 30 MM AUTOMATIC CANNON OPERATING MECHANISM

NIKOLA RADOVANOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia
e-mail: nikola.radovanovic@mod.gov.rs

ALEKSANDAR KARI
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia
e-mail: aleksandar.kari@mod.gov.rs

MILAN SIMONOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia
e-mail: milan.simonovic@mod.gov.rs

MARKO SIMONOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia
e-mail: marko.simonovic@mod.gov.rs

Abstract: *This paper presents a dynamic analysis of the operating mechanism of a 30 mm automatic cannon based on the gas-operated principle. By utilizing internal ballistics parameters obtained through the Drozdov method and the Bravin mathematical-mechanical model, a dynamic analysis of the bolt motion was performed. A MATLAB model of the 30 mm automatic cannon's bolt kinematics was developed, focusing on the dynamics during the recoil and counter-recoil cycles. This approach enabled the calculation of the theoretical rate of fire and illustrated the bolt velocity variations as a function of time. Finally, the obtained results were compared and validated against experimental data.*

Keywords: *30 mm automatic cannon, dynamic analysis, gas-operated mechanism, bolt kinematics, MATLAB modeling*

1. INTRODUCTION

A weapon based on the principle of gas operation is one that diverts part of the propellant gases through an opening in the barrel wall after the projectile passes that opening. This weapon uses the energy of the diverted gases to move the main assembly, which in turn drives all other parts of the weapon, while also accumulating energy in the return springs of the main assembly.

The device for diverting the propellant gases is most commonly called the gas regulator (gas piston system), and it consists of the following elements: the gas cylinder (often called the gas chamber), the working volume — the space into which the gas flows, ensuring the thermal process and the mechanical action of the gas — the gas port, a channel of various shapes through which the propellant gases pass from the barrel bore into the working volume of the gas regulator, the piston, which converts the gas action into kinetic energy, and the clearances and vent ports through which the propellant gases escape from the working volume, Fig. 1.

The problem of selecting the design parameters of the gas-operating system has given rise to a number of sub-variants of the piston-operated gas system: static, dynamic, static-dynamic, and others [1].

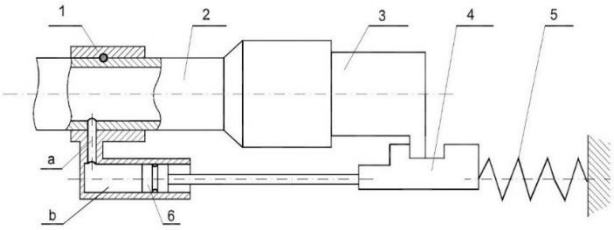


Figure 1. Diagram of automatic weapon based on the gas-operated principle
1-gas chamber, 2-barrel, 3-bolt, 4- bolt carrier, 5 – return spring, 6- piston, a- gas port, b – gas cylinder [2].

Figure 2 shows the characteristic variation of pressure in the gas regulator relative to the propellant gas pressure in the barrel, from the moment the projectile passes the gas port to the moment the outflow of the propellant gases ends.

The propellant gases behind the projectile, once its base passes the port, flow through the gas port and fill the cylinder containing the piston. The pressure of the diverted gas pushes the piston in the cylinder, so this work can be used to perform certain functions of the automatic weapon. This work is usually used to unlock the bolt or to actuate other automatic mechanisms.

Corresponding author: N. Radovanović;
E-mail: nikola.radovanovic@mod.gov.rs

DOI: 10.5937/OTEH26027R

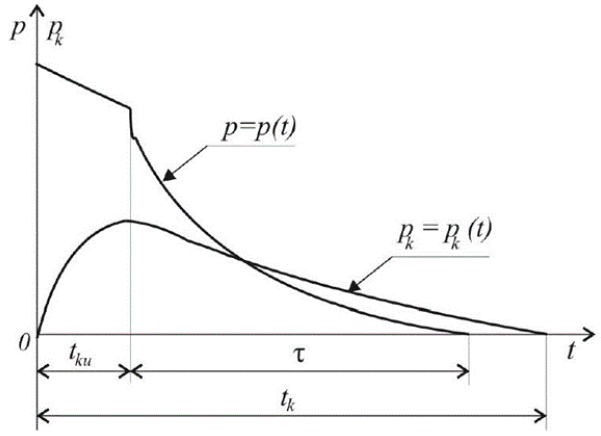


Figure 2. Pressure diagrams in the barrel and in the gas cylinder

The work of the diverted propellant gases can be used for various mechanisms, with the most common design solution being a cylinder with a piston and a piston rod for force transmission.

At the moment of firing, the bolt is firmly locked to the barrel. The force that propels the projectile and gases forward simultaneously acts on the case head in the opposite direction and produces recoil of the entire weapon.

When the projectile passes the gas port, the gases enter the gas port channel and generate a force on the gas piston.

The amount of gas required to move the piston is relatively small and does not significantly affect the muzzle velocity of the projectile.

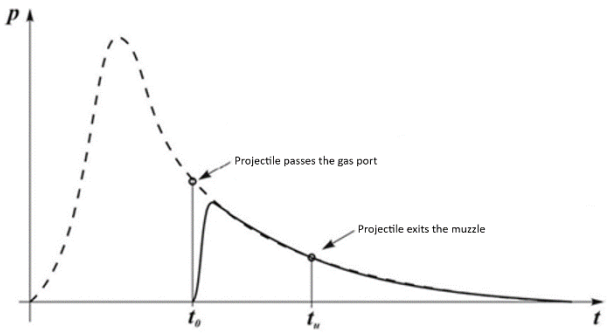


Figure 3. Propellant Gas Pressure Diagram

A piston with a larger effective area receives a greater impulse, and the magnitude of the impulse depends on the duration and magnitude of the gas pressure. This impulse depends directly on the size and shape of the port in the barrel wall. The magnitude of the piston energy depends on the piston mass, and for a given impulse, a piston of smaller mass has greater energy, since:

$$E = \frac{1}{2}mv^2 = \frac{1}{2}\frac{(mv)^2}{m} = \frac{1}{2}\frac{I^2}{m} \quad (1)$$

In accordance with this, the influencing factors that determine the kinetic energy of the piston are:

- the pressure in the barrel,
- the position and size of the gas port,
- the cylinder diameter (shape of the port),
- the piston area and its mass [1] [3] [4].

2. PROPELLANT GAS PRESSURE

The internal ballistics calculation was performed using the Drozdov method. The Drozdov method is used for weapons that employ a single type of propellant. These include small arms, guns, and howitzers when firing charges made of only one type of propellant. This method is based on the analytical solution of the system of internal ballistics equations [5].

For the adopted propellant data and the structural characteristics of the barrel of the 30 mm automatic cannon, an internal ballistics calculation was performed, followed by optimization aimed at achieving a higher initial velocity and lower propellant gas pressure, and the final law of propellant gas pressure variation along the barrel as a function of the projectile's travel distance was obtained, Figure 4.

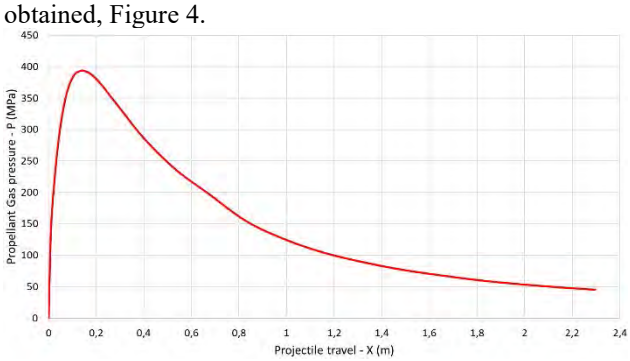


Figure 4. Optimized diagram of propellant gas pressure versus travel distance

Both domestic and foreign scientists have studied the examination of propellant gas pressures in gas regulators, with the aim of determining the law of variation *p=f(t)* and using that law to calculate the motion characteristics of the automatic weapon's main assembly. The problem of pressure in the gas regulator was addressed by Blagonravov A.A., Mamontov M.A., Bravin E.L., Paul Vieille, Jean Constantinet, Karl Becker, Hermann Scherer, Fiedner and Hutchings and others.

Bravin does not treat the process of gas outflow into the chamber as a thermodynamic process but instead offers an empirical expression for the propellant gas pressure in the gas chamber and on the piston of the main assembly, in order to represent as accurately as possible, the law of pressure variation of the gases entering the gas chamber [6].

As the basic factors influencing the law of propellant gas pressure variation on the piston, Bravin E.L.,[7] takes:

- the ratio of the cross-sectional areas of the gas port and the piston:

$$\sigma_0 = \frac{S_\phi}{S_n} \quad (2)$$

- the transverse load of the piston:

$$K_0 = \frac{M_0}{S_n} \quad (3)$$

In doing so, the following expression for the variation of gas pressure on the piston as a function of time is used:

$$p = p_\phi e^{-\frac{t}{b}}\left(1 - e^{-\alpha\frac{t}{b}}\right) \quad (4)$$

Where: $\alpha = n\frac{S_w M_w}{S_2^n}$; $b = \frac{i_0}{p_\phi}$; $i_0 = \frac{p_\phi + p_u}{2}t_n + \frac{p_u}{A}$

α – coefficient of useful action (efficiency);

b – Bravin's coefficient;

i_0 – unit impulse of the propellant gases in the barrel;

p_ϕ - the propellant gas pressure in the barrel bore at the moment the projectile passes the gas port;

p_u - the propellant gas pressure in the barrel bore at the moment the projectile exits the barrel;

$$A = 0.365\frac{\beta^2}{\beta - 0.5} * \frac{\vartheta_0}{L'} \quad (5)$$

$$\beta = C\frac{d}{\vartheta_0}\sqrt{\frac{p_\partial L' g}{\omega}} \quad (6)$$

C – an experimental coefficient equal to $C = 1.5$

t_n - the time from the moment the projectile passes the gas port until it exits the barrel;

M_n - the mass of the automatic weapon's driving assembly

i_0 - the unit impulse of the propellant gas pressure in the barrel bore during time t_g

Then, for the unit (specific) impulse in the gas-operating device, we obtain:

$$i_k = \int_0^t p_\phi e^{-\frac{t}{b}}\left(t - e^{-\alpha\frac{t}{b}}\right)dt =$$
$$= p_\phi b\left[\frac{\alpha}{1+\alpha} + \frac{\alpha}{1+\alpha}e^{-(1+\alpha)\frac{t}{b}} - e^{-\frac{t}{b}}\right] \quad (7)$$

For $t \to \infty$ $i_k = p_\phi b\frac{1}{1+\alpha}$ where $\frac{\alpha}{1+\alpha} = \eta$

The quantity η represents the efficiency coefficient of the gas-operating system.

For $\eta = 1$ the entire pressure impulse in the barrel bore would be transferred to the piston without losses.

The total unit impulse in the barrel bore is:

$$i_0 = \frac{p_\phi + p_\partial}{2}t_n + \frac{p_\partial}{A} \quad (8)$$

гд where A is from equation (4),

then: $p_\phi b = i_0$ and $b = \frac{i}{p_\phi}$

In doing so, when $\alpha \to \infty$ the expression for the pressure in the gas chamber takes the form:

$$p = p_\phi e^{-\frac{t}{b}}\left(1 - e^{-\alpha\frac{i}{e}}\right) \quad (9)$$

For $\alpha \to \infty$:

$$p = p_\phi e^{-\frac{t}{b}} \quad (10)$$

and will represent the limiting family of curves:

$$p = p_\phi e^{-\frac{t}{e}}\left(1 - e^{-\alpha\frac{t}{b}}\right) \quad (11)$$

The dependence: $p = p_\phi e^{-\frac{t}{b}}$ can approximately express the law of pressure variation in the barrel bore after the projectile passes the gas port until the end of the period.

The impulse of the propellant gas pressure force on the piston is determined based on the expression:

$$S_n i_k = S_n\left(\frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha}e^{-(1+\alpha)\frac{t}{b}} - e^{-\frac{t}{b}}\right)p_\phi b \quad (12)$$

But since $M_n V = S_n i_k$:

$$M_n V = S_n\left(\frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha}e^{-(1+\alpha)\frac{t}{b}} - e^{-\frac{t}{b}}\right)p_\phi b \quad (13)$$

For $t \to \infty$: $M_n V = S_n\frac{\alpha}{1+\alpha}$; for $M_n = const$ the expression for $M_n V$ allows the velocity of the main assembly to be determined at any moment in time.

By integrating the expression for velocity, we obtain the expression for the displacement of the main assembly:

$$X = \int_0^x dX = \int_0^t V dt = \frac{S_n b}{M_n}\left(\frac{\alpha}{1+\alpha}\cdot\frac{t}{b} - \frac{\alpha(2+\alpha)}{(1+\alpha)^2} - \frac{1}{(1+\alpha)^2}e^{-(1+\alpha)\frac{t}{b}} + e^{-\frac{t}{b}}\right)i_0 \quad (14)$$

For easier use of the expressions for determining p, V, X, coefficient tables have been formed:

$$K_p = e^{-\frac{t}{b}}\left(1 - e^{-\alpha\frac{t}{b}}\right) \quad (15)$$

$$K_v = \frac{\alpha}{1+\alpha} + \frac{1}{1+\alpha}e^{-(1+\alpha)\frac{t}{b}} - e^{-\frac{t}{b}} \quad (16)$$

$$K_x = \frac{\alpha}{1+\alpha}\cdot\frac{t}{b} - \frac{\alpha(2+\alpha)}{(1+\alpha)^2} - \frac{1}{(1+\alpha)^2}e^{-(1+\alpha)\frac{t}{b}} + e^{-\frac{t}{b}} \quad (17)$$

Then we obtain:

$$p = p_\phi K_p \quad (18)$$

$$V = \frac{S_n i_0}{M_n}K_v \quad (19)$$

$$X = \frac{S_n b i_0}{M_n}K_x = \frac{Vb}{K_v}K_x \quad (20)$$

To determine the quantity $\eta_k = \frac{\alpha}{1+\alpha}\eta_0 v_0 v_g v_{g0}$ dependency tables have been formed

$$\eta_0 = f(\sigma_\Pi, \sigma_\Delta)$$
$$v_0 = f(\sigma_\Pi, \sigma_0)$$
$$v_g = f(\sigma_\Pi, \sigma_g)$$
$$v_{g0} = f(\sigma_0, \sigma_g)$$

The stated tables for determining the impulse efficiency coefficients can be used to calculate the pressure and motion characteristics of the automatic weapon's main assembly during its free travel, both according to the method of Mamontov M.A. and according to the method of Bravin E.L. [6] [7].

For the purpose of simulating the pressure in the gas

cylinder, a software solution was developed in MATLAB, whose algorithm is shown in Figure 5.

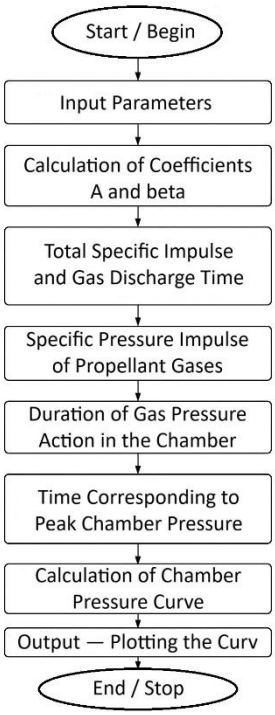


Figure 5. Algorithm of the software solution of Bravin's method

The software solution also enables the calculation of the gas-operating system parameters by simply varying the adopted design parameters. The input quantities used are the barrel caliber, projectile mass, propellant charge mass, bolt mass, initial projectile velocity, maximum propellant gas pressure, the propellant gas pressure at the muzzle and the propellant gas pressure at the gas port location obtained from the internal ballistics calculation, the diameter of the gas port, the piston diameter, the diametral clearance between the piston and the chamber, the initial chamber volume, the reduced barrel length, the projectile travel time from the gas port to the muzzle, the initial return spring force, the return spring stiffness, and the sliding friction coefficient. As output, a diagram of pressure variation in the barrel and in the gas regulator chamber is obtained [8].

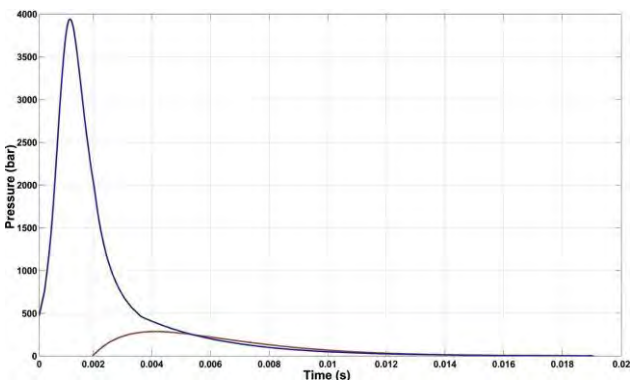


Figure 6. Diagram of pressure variation in the barrel and in the gas regulator chamber.

The obtained pressure variation diagram (Figure 6) correlates with the theoretical diagram shown in Figure 4. The pressure in the chamber begins to rise from the moment t_{k0} = 0.0020 *s*, i.e., when the projectile passes the gas port. The pressure in the barrel is dropping sharply, and at the moment the projectile is at the gas port it equals 2105 *bar*. Based on the diagram, it can be observed that the pressure in the chamber rises more slowly than the pressure in the barrel and is maintained long enough to ensure reliable operation of the automatics even when the barrel pressure is lower. It reaches its maximum value p_{km} = 284 *bar*. The total duration of the propellant gas action in the chamber is t_k = 0.0188 *s*. The pressure curves in the barrel and chamber nearly coincide and continue to decrease during the period of after-effect of the propellant gases, reaching zero after 0.02 *s*.

3. DYNAMIC ANALYSIS

3.1. Work Cycle Diagram

The work cycle diagram of the automatic mechanism represents a graphical or tabular depiction of the interconnected operating phases that occur during a single shot, i.e., during the entire operating cycle of the weapon. The cycle diagram shows the sequence, duration, and overlap of individual processes. Essentially, the cycle diagram is a schematic model of the working process, serving as a basis for studying, designing, and improving the weapon.

For the purpose of researching the operation of the automatics of the 30 mm automatic cannon, a high-speed camera was used, the *Phantom v1212.* The use of a high-speed camera represents one of the key methods for the nonlinear analysis of part motion in real time. Given that the bolt's operating cycle takes place within fractions of a second, standard measurement methods often cannot capture the critical moments that define phase changes. Based on the high-resolution recordings, a work cycle diagram of the 30 mm automatic cannon was constructed (Table 1). The masses of the automatic mechanism's elements are shown in Table 2.

Table 1. Work cycle diagram of the 30 mm automatic cannon

Phases	No.	Motion of parts	Travel distance (*mm*)	Time (*s*)	Mass (*kg*)
Recoil	1	Unlocking	0-35	0.01	m1+m7+m6
	2	Joint motion	35-350	0.02	m3+m4+m6+m7
	3	Impact on case ejector	350		
	4	Case ejection	350-400	0.006	m3+m6+m7
	5	Impact on buffer	400		
	6	Buffer compression and end of recoil	400-420	0.004	m1+m2+m6+m7
Counter-recoil Phases	1	Joint motion	420-390	0.05	m1+m2+m6+m7
	2	Impact of bolt on case head	390		
	3	Joint motion	390-35	0.04	m3+m5+m6+m7
	4	Impact of bolt on rear of barrel	35		

Table 2. Masses of the automatic mechanism's elements

Bolt carrier	m1
Bolt	m2
Bolt assembly	m3
Case	m4
Cartridge	m5
Piston	m6
Return spring	m7

4. BOLT MOTION KINEMATICS

In the paper, a model of the kinematics of the motion of the bolt of the 30 mm automatic cannon was developed using code, focusing on the multi-phase dynamics during the recoil and counter-recoil cycle. The model uses the real propellant gas pressure profile in the gas chamber as input to the dynamic model of the motion of the bolt assembly.

The simulation is based on the numerical solution of a system of second-order differential equations describing the motion of the bolt as a function of time. The complete cycle is divided into seven discrete phases, where each phase has a defined displacement length, the action of the propellant gas force, and the influence of the buffer. For each phase, the ODE system is solved separately using the function *ode45*, using an *event* -function for automatic detection of the end of the phase. This achieves a modular and physically consistent simulation [9].

The input pressure *pk(t)* is loaded from a *.mat* file obtained from the previous calculation of the pressure in the gas cylinder and is interpolated in time. The dynamics of the bolt include the effect of the spring, friction, and buffer, with parameters defined according to the technical characteristics of the system. The recoiling mass dynamically changes in the corresponding phases depending on the mechanical state (presence of the cartridge, case, etc.).

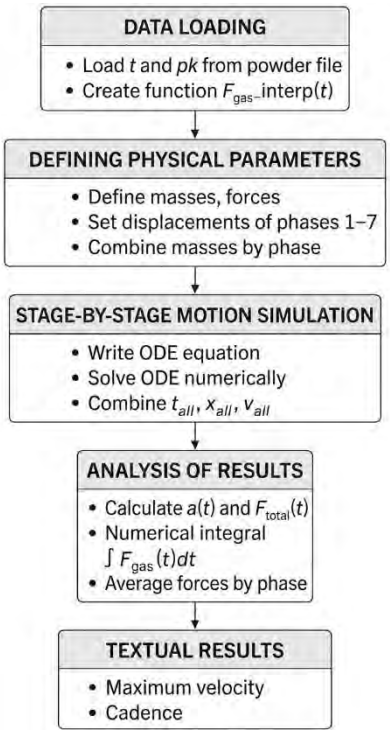


Figure 7. Algorithm of the bolt motion kinematics

Based on the numerical results, the key values of the system were calculated, including the maximum bolt velocity and the rate of fire.

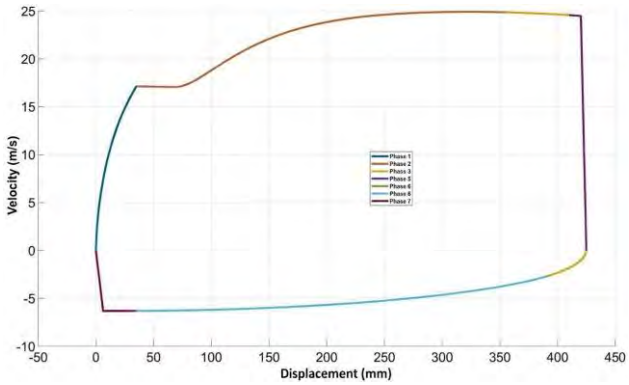


Figure 8. Bolt velocity as a function of position (V=f(x)).

The first phase represents the unlocking of the bolt. In this phase, the propellant gases exert the greatest influence on the system. The velocity of the bolt carrier increases sharply, and the diagram shows that the maximum acceleration occurs precisely here. This phase is crucial for the initial actuation of the mechanism and confirms that the gas profile correctly transfers energy to the piston.

After the initial acceleration and unlocking, the second phase involves extraction of the case, during which the entire bolt assembly together with the case joins the recoiling mass. Based on the diagram in Figure 9 and 10, a brief interval of constant bolt-assembly velocity can be observed at the very beginning of the second phase. This occurrence is a consequence of the sudden increase in inertial mass. Namely, the propellant gas pressure at that moment acts on a mass approximately three times greater than in the initial phase of motion. Thereafter, the velocity remains high but stable, while the gas force gradually decreases. The second phase ensures continuity of motion and shows that the system has enough energy to overcome the inertia of the bolt mass.

In the third phase, the gas impulse ceases, and the bolt continues its motion owing to the accumulated energy. The velocity decreases, and the average force is smaller. In this phase, the ejection of the case occurs which additionally slows the system down and represents a real mechanical effect in the operation of the weapon.

In the fourth phase, the bolt strikes the buffer. This phase clearly shows the dissipation of energy. The bolt velocity decreases, and the buffer absorbs the impact, which is seen through the sudden drop in force.

The bolt then comes to a temporary stop, which is reflected in the negative velocity values. The fifth phase provides a brief pause in the cycle and enables mechanical preparation for the return stroke. The diagram shows that the forces in this phase are negative, which is a consequence of the return action of the spring.

The longest phase of the return motion is the motion of the bolt together with the cartridge. The velocity is negative and stable, and the average forces indicate the dominant influence of spring and friction. This phase is crucial for chambering the next cartridge and preparing for the

following cycle.

The final phase represents the locking of the bolt. The bolt completes its return motion and assumes its initial position. The velocity decreases toward zero. This phase ensures the stabilization of the system and prepares it for the next firing cycle.

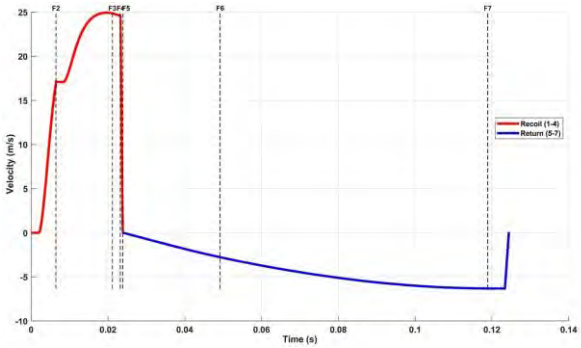


Figure 9. Variation of bolt velocity as a function of time (V=f(s)).

Analysis of the diagrams and tables confirms that the bolt dynamics proceed in accordance with the expected physical principles. The obtained rate of fire of 482 *rounds/min* shows a relative deviation of ~5%, compared to the know value. The diagram of bolt velocity variation, with a maximum value of 24.92 *m/s*, is free of sudden oscillations that could lead to dynamic overloads, which indicates stable operation of the system.

5. CONCLUSION

Based on the available data on the 30 mm automatic cannon and the 30×173 mm cartridge, dynamic analysis of 30 mm automatic cannon operating mechanism was performed.

The internal ballistics calculation was performed using the Drozdov method, and after optimization, the maximum propellant gas pressure in the barrel of the 30 mm automatic cannon was obtained as $P_m=394\ MPa$, while the muzzle velocity of the projectile is $V_0=1104\ m/s$.

Starting from the conditions of the set task, the method of theoretical examination of propellant gas pressures in gas-operating devices according to Bravin's method was considered. The resulting pressure curve in the gas cylinder fully correlates with the theoretical and experimental work in this field. The maximum pressure value in the chamber is $p_{km}=283.76\ bar$, the duration of the propellant gas action in the chamber is $t_k=0.0188\ s$, and the time corresponding to the maximum pressure in the chamber is $t_{km}=0.0021\ s$.

Based on the constructed work cycle diagram of the 30 mm automatic cannon, a simulation of the bolt motion through seven phases with variable masses was performed, using numerical integration by means of the *ode45* method. The previously obtained pressure curve in the chamber was used as the input data. The displacement and velocity diagrams by phase show that the greatest acceleration occurs in the first phases (1–2), when the gases act directly on the piston and the bolt carrier. In phase 3, the force and velocity decrease due to the resistance of case extraction, while phase 4 clearly shows the dissipation of energy through the buffer. The return phases (5–7) are characterized by negative velocity, confirming the correct return of the bolt to its initial position. This distribution confirms that the model successfully captures the transition between recoil and counter-recoil, and that the buffer plays a key role in stabilizing the system. A rate of fire of 482 rounds per minute was obtained, which deviates from the experimental value by ~5%, confirming that the simulation faithfully describes the real operation of the bolt. Each phase has a clearly defined dynamic, ranging from the initial acceleration under the gas impulse, through energy dissipation in the buffer, to the return motion and stabilization by locking of the bolt. The obtained results are consistent with the expected mechanical behavior and represent a reliable basis for further optimization of the design.

Further improvements to the simulation could be achieved by introducing spring nonlinearities for a more realistic representation of the elastic forces, as well as by applying different loss coefficients in the gas system. In addition, instead of idealized viscous friction, a real friction model that better reflects the actual mechanical operating conditions should be introduced. Finally, extending the simulation to a multi-cycle regime, instead of a single-cycle one, would enable the analysis of the system's durability, stability, and reliability under conditions of intensive use.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] Petrović M.: Mehanika automatskog oružanja (Mechanics of Automatic Weapons), Media Center "Odbrana", Belgrade, 2009.

[2] V. D. Tien, M. Macko, S. Procházka, V. V. Bien, Mathematical Model of a Gas-Operated Machine Gun, Advances in Military Technology, vol. 17, no. 1, pp. 63–77, 2022.

[3] Petrović M., Kari A.: Zbirka zadataka iz mehanike naoružanja 2 (Collection of Problems in Weapon Mechanics 2), Military Academy, Belgrade, 2010.

[4] Ristić Z.: Zbirka zadataka iz mehanike naoružanja (Collection of Problems in Weapon Mechanics), VIZ – Printing and Reproduction Section, Belgrade, 2004.

[5] Cvetković M.: Unutrašnja balistika (Internal Ballistics), Military Publishing Institute, Belgrade, 1998.

[6] Tančić Lj.: Zbirka zadataka iz unutrašnje balistike (Collection of Problems in Internal Ballistics), Military Academy, Belgrade, 1997.

[7] Bravin E.L., Strelkovoe pušečnoe vooruženie samoletov, Moskva, 1941.

[8] Radovanovic N., Kari A.: Analysis of Construction Parameters of Gas Operated Weapons, OTEH, pp. 116–121, Belgrade, 2018.

[9] P. D. Nguyen, B. V. Vo, M. N. Nguyen, and H. T. Tran, "Dynamics of Automatic Firing System Type Gas-Operated Weapon," Problemy Mechatroniki: uzbrojenie, lotnictwo, inżynieria bezpieczeństwa, vol. 14, no. 4, pp. 9–22, 2023.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTERNAL BALLISTIC DESIGN OF A 30 x 173 mm AUTOMATIC CANNON USING THE ANALYTIC METHOD

MARIJA MILOVANOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia, e-mail: marija.milovanovic@mod.gov.rs

NIKOLA RADOVANOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia, e-mail: nikola.radovanovic@mod.gov.rs

STEFAN JOVANOVIĆ
Military Technical Institute, Ratka Resanovića 1, 11000 Belgrade, Serbia, e-mail: stefan.jovanovic@mod.gov.rs

Abstract: *This paper presents the internal ballistic design of a 30 x 173 mm automatic cannon. The initial section provides general observations on 30 mm automatic cannons, including the domestic M12 automatic model. The internal ballistic calculation was conducted using the Drozdov method. Through the variation of ballistic parameters, an optimal variant was selected. Furthermore, experimentally determined internal ballistic (IB) data are presented, and a comparative analysis between the obtained results and theoretical models has been performed.*

Keywords: *Internal Ballistics, Automatic Cannon, 30 x 173 mm Caliber, Drozdov Method, Ballistic Design*

1. INTRODUCTION

Modern armed conflicts and the rapid evolution of land, naval, and air combat doctrines have imposed increasingly demanding requirements on small- and medium-caliber automatic weapon systems. Within the category of automatic cannons intended for installation on infantry fighting vehicles (IFVs), tactical aircraft, naval platforms, and air defense systems, the 30 mm caliber has become the optimum compromise between ballistic performance, system mass, and recoil-force management [1].

The transition from 20 mm and 25 mm weapon systems to the 30 mm caliber has primarily been driven by the increasing effectiveness of modern armor protection systems and the demand for greater terminal effectiveness against a wide spectrum of targets. The larger projectile volume available in the 30 mm caliber enables the employment of highly efficient Armor-Piercing Fin-Stabilized Discarding Sabot (APFSDS) projectiles, as well as advanced programmable airburst ammunition (ABM), which has become essential for engaging concealed targets and countering unmanned aerial systems (C-UAS).

1.1. 30 mm Automatic Cannons

Modern single-barrel 30 mm automatic cannons are generally equipped with dual-feed ammunition systems, allowing rapid selection between two different ammunition types and thereby significantly improving combat effectiveness against various ground, aerial, and naval targets. Typically, one ammunition belt is loaded with High-Explosive (HE), High-Explosive Tracer (HET), or High-Explosive Incendiary Tracer (HEIT) projectiles, whereas the second belt contains Armor-Piercing (AP), Armor-Piercing Tracer (APT), or Armor-Piercing Incendiary Tracer (APIT) ammunition. Fragmentation ammunition is additionally employed against enemy personnel and lightly protected targets.

The 30 × 173 mm caliber is widely adopted by NATO member states and countries operating NATO-compatible weapon systems, whereas the 30 × 165 mm caliber is predominantly used by the armed forces of the Russian Federation.

Among the most widely deployed 30 × 173 mm automatic cannons are:

- Oerlikon KCE (Switzerland)
- GAU-8/A Avenger (United States)
- Bushmaster II Mk44 (United States)
- Rheinmetall Mk30-2 (Germany)

The primary armament of numerous United States armored combat vehicles is the Bushmaster II Mk44 automatic cannon (Figure 1). The Bushmaster II Mk44 is a 30 × 173 mm dual-feed automatic cannon with a cyclic rate of fire of approximately 200 rounds per minute and an effective firing range of up to 3,000 m. It is recognized for its exceptional reliability and operational performance on modern armored vehicles and naval platforms.

Figure 1. General view of the Bushmaster II Mk44

The domestically developed M12 30 mm automatic cannon, shown in Figure 2, is designed for engaging

Corresponding author: M. Milovanovic;
E-mail: marija.milovanovic@mod.gov.rs
DOI: 10.5937/OTEH26028M

personnel, armored and unarmored ground targets, as well as aerial threats. The system provides an effective firing range of at least 3,000 m against ground and surface targets and at least 1,500 m against aerial targets.

Table 1. Technical characteristics of 30 mm cannon M12

Weapon type:	single-barrel automatic cannon
Caliber:	30 × 173 mm
Automatic operating principle:	gas-operated with barrel recoil
Length:	3255 mm
Barrel length	2460 mm
Barrel rifling:	progressive, from 2.5° to 8.5°
Number of grooves:	16
Cannon mass:	195 kg
Rate of fire:	from 350 rds/min to 430 rds/min
Initial velocity of projectile:	(1100 ± 15) m/s
Firing mechanism:	mechanical
Recoil force:	from 15 kN to 20 kN
Operating temperature range:	from –30 °C to +50 °C
Barrel ballistic life:	up to 6000 rounds

Figure 2. External view of the M12 30 mm automatic cannon

1.2. 30 x 173 mm Ammunition

Ammunition for 30 × 173 mm automatic cannons is generally manufactured in three primary variants:

- High-Explosive (HE);
- Armor-Piercing (AP);
- Armor-Piercing Fin-Stabilized Discarding Sabot (APFSDS).

HE projectiles are available in tracer configurations, whereas AP projectiles may incorporate tracer and incendiary elements to improve target indication and post-penetration effectiveness. Table 2 summarizes the principal technical characteristics of representative 30 × 173 mm cartridges. Figure 3 illustrates the general appearance of several cartridge types, while Figure 4 presents a cartridge fitted with an APFSDS projectile.

Table 2. Technical characteristics of 30x173 mm cartridges

	Mk238 High-explosive fragmentation with tracer	**Mk239** Training round with tracer	**Mk268** Armor-piercing with tracer	**Mk310** High-explosive
Round length	290 mm	290 mm	290 mm	290 mm
Round mass	670 g	675 g	725 g	713 g
Mass of projectile	362 g	372 g	235 g	423 g
Propellant charge mass	153 g	153 g	176 g	150 g
Fuze	Impact, instantaneous action, with self-destruct	Inert	No fuze	Programmable fuze
Projectile velocity	1050 m/s	1050 m/s	1050 m/s	980 m/s
Maximum pressure	3450 bar	3450 bar	3450 bar	3600 bar
Dispersion	0,5 mr	0,5 mr	0,5 mr	0,5 mr

Figure 3. –General appearance of several cartridge types

Figure 4. – Cartridge fitted with an APFSDS projectile

2. INTERNAL BALLISTIC ANALYSIS

The internal ballistic analysis was performed using the Drozdov analytical method, which is applicable to weapon systems employing a single type of propellant. This method is commonly used in the internal ballistic design of small arms, automatic cannons, field guns, and howitzers firing ammunition loaded with a single propellant type. The Drozdov method is based on the analytical solution of the governing equations of internal ballistics [2].

The mathematical description of projectile motion during the firing cycle is obtained by solving the system of equations over several characteristic periods of the firing process. The exception is the time function, which is evaluated continuously throughout the entire firing cycle.

Because the actual internal ballistic process is highly complex and several influencing parameters cannot be determined precisely, the Drozdov method introduces a number of simplifying assumptions that enable an analytical solution of the governing equations. These assumptions are summarized as follows:

- Propellant combustion follows the geometric burning law.
- Combustion takes place under the average pressure of the propellant gases.
- The burning-rate law is assumed to be linear with pressure.
- Secondary work performed by the propellant gases is assumed to be proportional to the translational work of the projectile and is accounted for through the fictitious coefficient, φ.

- Heat losses are considered indirectly by reducing the specific energy of the propellant gases or by increasing the thermodynamic exponent θ.
- The composition of the combustion products remains constant throughout the firing process; therefore, both the specific energy of the propellant gases (fb) and the co-volume (α) are assumed constant.
- The work required for engraving the rotating band into the rifling is neglected. The projectile is assumed to remain stationary until the propellant-gas pressure reaches the forcing pressure (p_0).
- Projectile engraving is assumed to occur instantaneously rather than progressively.
- Projectile motion is analyzed only until the projectile base exits the muzzle.
- Elastic deformation of the barrel, gas leakage between the rotating band and the bore, and aerodynamic resistance inside the barrel are neglected.
- In gas-operated weapon systems, propellant-gas extraction begins simultaneously with projectile motion and is assumed to be steady, without expulsion of unburned propellant particles.

All constant quantities appearing in the governing equations are defined as input parameters whose numerical values are known in advance. These parameters are grouped into three categories:

- barrel characteristics: S_c, W_0, X_u;
- projectile characteristics: m, p_0, φ and
- propellant characteristics: m_b, f_b, α, u_{z0}, ρ_b, θ, $2r_0$, κ, λ, μ, (κl, λl).

The Drozdov method is originally formulated for regressive-burning propellant grains. When progressive-burning grains, such as seven-perforated cylindrical propellant grains, are employed, the grain web thickness is multiplied by a factor of 10/7, while the grain-shape coefficients are determined using the equivalent single-perforated cylindrical grain geometry [3] [4].

The firing cycle is divided into three characteristic periods, Preliminary, First and Second. During the preliminary period, the projectile remains stationary while the propellant burns within a constant chamber volume. Consequently, only the state variables at the end of this period are required, since they constitute the initial conditions for the subsequent calculation. The first period extends from projectile motion initiation until complete combustion of the propellant charge. During this stage, the maximum chamber pressure is reached. The second period begins after complete propellant combustion and lasts until the projectile exits the muzzle. During this interval, the projectile continues to accelerate solely under the action of the expanding propellant gases. A third period, commonly referred to as the after-effect period, describes the influence of the propellant gases after projectile exit and is not included in the present internal ballistic calculation [3].

2.1. SOFTWERE IMPLEMENTATION

The computational algorithm is organized according to the characteristic periods of the firing process defined by the Drozdov analytical method. Initially, the program reads all input parameters from the ULD100 input file. These parameters are subsequently used to calculate the required constants and the characteristic quantities corresponding to the preliminary period. For the first period, the initial conditions, computational interval, and integration step are defined. Based on the selected interval and the number of calculation points, the computational step size is determined. A DO loop is then used to calculate the following internal ballistic parameters at each computational step: chamber pressure, *p(x)*; projectile travel, *X(x)*; projectile velocity, *V(x)*; elapsed time, *t(x)*. During this period, the program automatically determines the maximum chamber pressure. The second period is solved using the same computational procedure. Appropriate initial conditions, computational interval, and step size are defined, after which the following quantities are calculated for every integration step: chamber pressure, *p(X)*; projectile velocity, *V(X)*; elapsed time, *t(X)*. The ULD100 file contains all input parameters separated by commas and organized into four rows. During execution, the program automatically generates the output file I100, in which all calculated ballistic parameters are stored through successive WRITE commands.

The developed computational procedure is universal and can therefore be applied to any weapon system employing a single type of propellant [5].

The internal ballistic analysis presented in this study was carried out for the 30 mm automatic cannon. This weapon employs NC-52 seven-perforated cylindrical propellant grains, illustrated in Figure 5, consisting of one central perforation surrounded by six equally spaced peripheral perforations arranged in a regular hexagonal pattern.

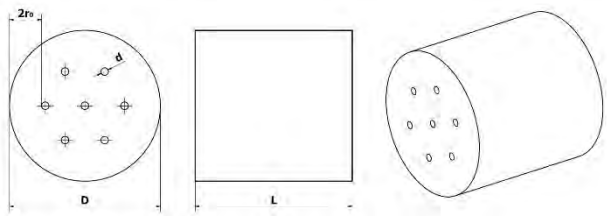


Figure 5. General appearance of the NC-52 propellant grain.

The data required for the internal ballistic (IB) calculation are:

- Powder chamber volume:

$$W_0 = \frac{m_b}{\rho} \quad (1)$$

$$W_0 = \frac{0{,}161}{700} = 0{,}00023$$

ρ for high-power artillery weapons, from 650 to 780 kg/m^3

- Barrel cross-sectional area:

$$S_c = n_s \cdot d^2 \quad (2)$$

$$S_c = 0{,}82 \cdot 0{,}030^2 = 0{,}000738\ m^2$$

- Total travel of the projectile in the barrel:

$$X_u = 2296\ m$$

- Fictitiousness (form) coefficient:

$$\varphi = K + \frac{1}{3} \cdot \frac{m_b}{m} = 1{,}03 + \frac{1}{3} \cdot \frac{0{,}161}{0{,}357} = 1{,}1767505 \quad (3)$$

Where

K is approximately 1.03 for high-power guns

- Projectile mass:

$$m = 0{,}357\ kg$$

- Propellant mass:

$$m_b = 0{,}161\ kg$$

- Density of the propellant charge:

$$\rho_b = 1600\frac{kg}{m^3}$$

- Covolume of the propellant gases:

$$\alpha = \left(1456 - \frac{0{,}524}{4{,}187}\cdot Q_w\right)\cdot 10^{-6} \qquad (4)$$

$$\alpha = \left(1456 - \frac{0{,}524}{4{,}187}\cdot 3300\right)\cdot 10^{-6} = 0{,}00104300741$$

Where:

Q_w = 3200–3400 J/kg – heat quantity of the propellant at constant volume.

- Ratio of specific heats minus 1:

$$\theta = 0{,}269$$

- Half-thickness of the propellant grain:

$$r_0 = 0{,}3$$

- Shape coefficients:

For seven-perforated propellant, the grain thickness is multiplied by a factor of 10/7, and the shape coefficients are calculated as for a cylindrical single-perforated grain.

$$\kappa = 1 + \beta\,, \qquad \lambda = -\frac{\beta}{1+\beta}\,, \qquad \mu = 0 \qquad (5)$$

$$\kappa_1 = \kappa\left(1 - \frac{\mu}{2}\right), \qquad \lambda_1 = \frac{1}{\kappa_1} - 1 \qquad (6)$$

$$\beta = \frac{2r_0}{2c} = 0{,}3296703$$

$$\kappa_1 = 1{,}32967\,, \qquad \lambda_1 = -0{,}247934$$

- Specific work of the propellant gases:

$$f_b = \left(6546 + \frac{3.4001}{4.187}\cdot Q_w\right)\cdot 10^2 \qquad (7)$$

$$f_b = \left(6546 + \frac{3.4001}{4.187}\cdot 3300\right)\cdot 10^2 = 922580$$

- Forcing pressure

$$p_0 = 40\cdot 10^6$$

- Unit burning rate of the propellant:

$$u_{z0} = \left(-76.6 + \frac{0.179}{4.187}\cdot Q_w\right)\cdot 10^{-11} \qquad (8)$$

$$u_{z0} = \left(-76.6 + \frac{0.179}{4.187}\cdot 3300\right)\cdot 10^{-11}$$

$$= 6{,}45\cdot 10^{-10}\ \frac{m}{sPa}$$

Table 3. Input data for the internal ballistic calculation

Name	Symbol	Value	Unit
Initial powder chamber volume	W_0	0,00023	m^3
Barrel cross-sectional area	S_c	0,000738	m^2
Total travel of the projectile	x_u	2,296	m
Fictitiousness coefficient	φ	1.17675	/
Projectile mass	m	0,357	kg
Propellant mass	m_b	0.161	kg
Bulk density	ρ_b	1600	kg/m^3
Covolume of the propellant gases	α	0,00104301	m^3/kg
Ratio of specific heats minus 1	Θ	0,269	/
Half-thickness of the propellant grain	r_0	0,0003	m
Shape coefficient	κ_1	1,32967	/
Shape coefficient	λ_1	- 0,247934	/
Specific work of the propellant gases	f_b	922580	J/kg
Forcing pressure	p_0	$40\cdot 10^6$	Pa
Unit burning rate of the propellant	u_{z0}	$0{,}000645\cdot 10^{-6}$	m/sPa

The graphical interpretation of the obtained results was done using the software package *Microsoft Excel,* showing the dependence of propellant gas pressure (*P*) on time (*t*) and travel (*X*), as well as the change in projectile velocity (*V*) with time (*t*)), is shown in Figures 6, 7, and 8.

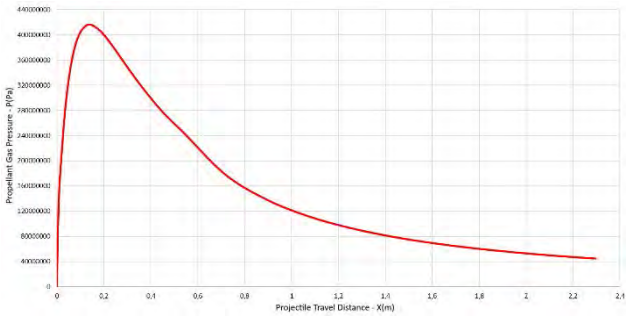


Figure 6. Change in propellant gas pressure as a function of travel

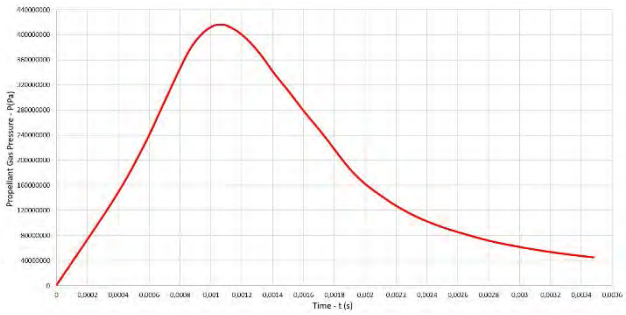


Figure 7. Change in propellant gas pressure as a function of time

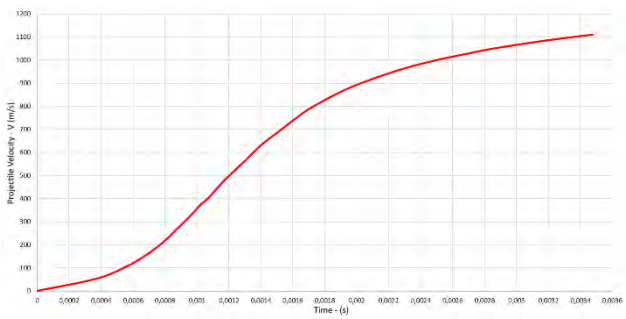


Figure 8. Change in projectile velocity with time

From the presented IB calculation, it can be seen that the maximum propellant gas pressure in the barrel of the 30 mm M12 cannon is P_m=*416 MPa*, the time elapsed from firing until the projectile leaves the muzzle is t_u=*0,00348*

s, the distance traveled by the projectile is x_u=*2,3 m*, while the projectile velocity at the muzzle is V_0=*1110 m/s*.

3. PARAMETRIC ANALYSIS OF IB PARAMETARS

The ballistic characteristics of the propellant are those characteristics on which the law governing the development of propellant gas pressure depends. In order to determine the influence of individual parameters on the magnitude of the propellant gas pressure in the barrel and on projectile velocity, they will be varied within a range of ±2.5% and ±5%. Parameters will be varied one at a time, while the remaining parameters retain their baseline value. In this way, an optimal solution can be reached. The goal is to find parameter values such that the propellant gas pressure in the barrel is as low as possible while the initial projectile velocity is as high as possible. The parameters that are varied are:

- the covolume of the propellant gases α,
- the specific work of the propellant gases f_b,
- the unit burning rate of the propellant u_{z0}.

3.1. Influence of the propellant gas covolume (α)

The covolume of the propellant gases (α) represents the smallest volume to which the propellant gases produced by burning 1 *kg* of propellant can be compressed. It is, in essence, the volume of the propellant gas molecules [3] [4].

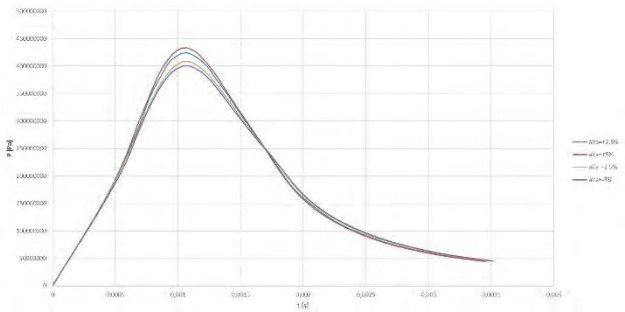


Figure 9. Influence of the propellant gas covolume on the propellant gas pressure in the barrel

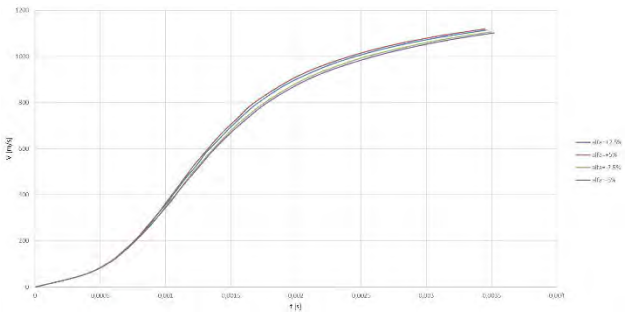


Figure 10. Influence of the covolume on projectile velocity in the barrel

Based on the diagrams in Figures 9 and 10, it can be concluded that as the propellant gas covolume α increases, the propellant gas pressure in the barrel and the projectile velocity increase, and vice versa with a decrease.

3.2. Influence of the unit burning rate (u_{z0})

The unit burning rate (u_{z0}) is a constant from the burning law of the propellant grain:

$$u_z = u_{zo}p \tag{9}$$

For nitrocellulose propellants it amounts to (5.5–8)10^{-10} *m/sPa*, while for nitroglycerine propellants it is (4–12)10^{-10} *m/sPa*.

The ignition velocity, i.e., the rate of flame propagation over the surface of the propellant grain, is 2–4 *mm/s* for smokeless propellants and 1–3 *m/s* for black powder. The burning rate, i.e., the rate of flame penetration into the depth of the propellant grain, is 1–1.5 *mm/s* for smokeless propellants and about 1 *cm/s* for black powder [3] [4].

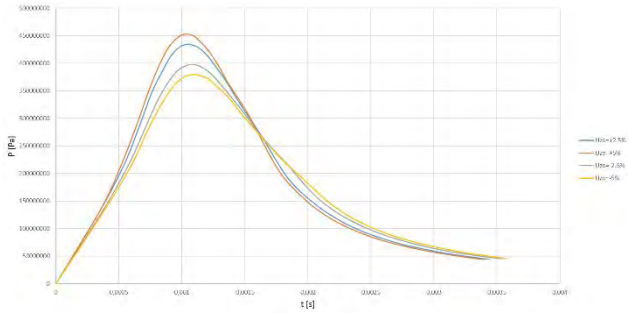


Figure 11. Influence of the unit burning rate on the propellant gas pressure in the barrel

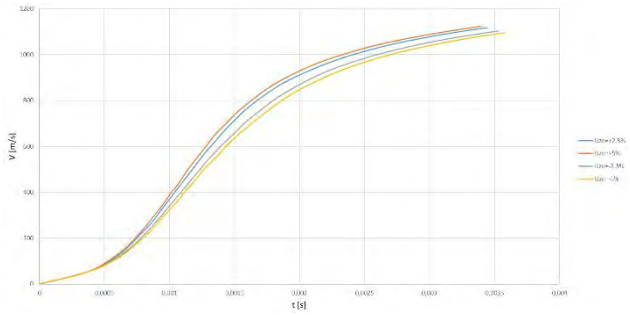


Figure 12. Influence of the unit burning rate on projectile velocity in the barrel

Figure 11 shows the change in propellant gas pressure with travel due to a change in the unit burning rate, while Figure 14 shows the change in projectile velocity with travel. The unit burning rate has the same effect as the propellant gas covolume, i.e., increasing its nominal value increases the propellant gas pressure and the projectile velocity.

3.3. Influence of the specific work of the propellant gases (f_b)

The specific work of the propellant gases (f_b) represents the work that could be performed by the propellant gases produced by burning 1 *kg* of propellant if cooled to 0˚*C* and expand at atmospheric pressure. It ranges within 1±0.25 *MJ/kg*. It is determined experimentally by burning the propellant in a manometric bomb, or by thermochemical calculation using the expression:

$$f_b = MRT_m \tag{10}$$

Where:
M - total number of moles of the gas mixture

R - universal gas constant, and
T_m - maximum temperature of the propellant gases.

The gas constant is determined from the ideal gas equation of state, for a molar volume of W_m=0.02241 m^3/mol, atmospheric pressure p_a=101334 Pa and temperature T_a=273,15 K:

$$R = \frac{p_a W_m}{T_m} = 8.314 \frac{J}{mol \cdot K} \quad (11)$$

Figures 13 and 14 show the change in propellant gas pressure and the change in projectile velocity in the barrel due to the influence of the specific work of the propellant gases (f_b).

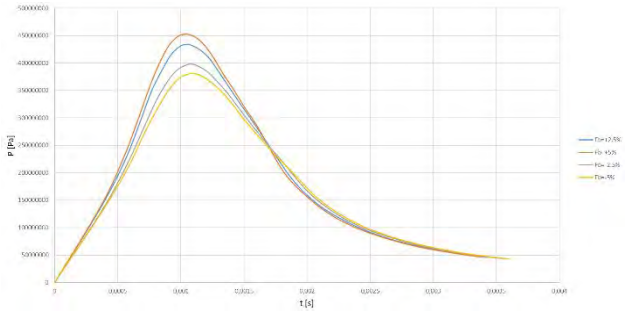


Figure 13. Influence of the specific work of the propellant gases on the propellant gas pressure in the barrel

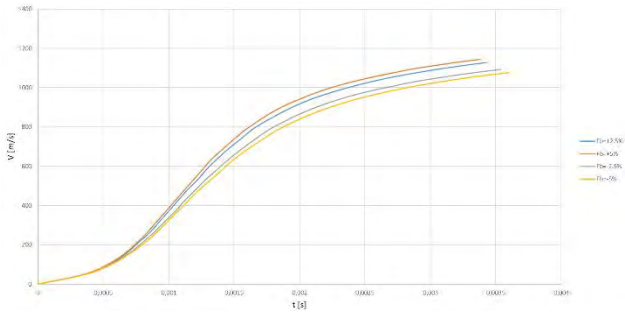


Figure 14. Influence of the specific work of the propellant gases on projectile velocity in the barrel

Reducing the specific work of the propellant gases reduces the maximum propellant gas pressure in the barrel and the initial velocity. Compared with the propellant gas covolume and the unit burning rate, for the same percentage change relative to the nominal value, the specific work of the propellant gases has the greatest influence on the change in pressure and velocity.

3.4. Optimal Ballistic Solution

The automatic 30 mm M12 cannon has a maximum barrel pressure of P_m=420 MPa, and an initial projectile velocity of V_0=1100 m/s. The task is to find a solution that satisfies the relations P_m<420 MPa, V_0>1100 m/s.

Table 4 gives the values of the maximum propellant gas pressure and the initial projectile velocity when varying the selected parameters.

Table 4. Values of Pm and V_0 for variation of the parameters

	-2.5%P_m [MPa]	-2.5%V_0 [m/s]	-5%P_m [MPa]	-5%V_0 [m/s]	P_m [MPa]	V_0 [m/s]	+2.5%P_m [MPa]	+2.5%V_0 [m/s]	+5%P_m [MPa]	+5%V_0 [m/s]
α	408	1106	401	1102	416	1110	424	1115	433	1119
u_{z0}	398	1103	380	1094	416	1110	435	1118	454	1124
f_b	398	1094	381	1076	416	1110	434	1127	453	1144

The propellant gas covolume gives the optimal solution when its value is increased by 2.5%. Comparing the specific work of the propellant gases with the unit burning rate, it is observed that for the same maximum pressure, the unit burning rate yields a higher initial velocity.

The diagrams in Figures 15, 16, and 17 show the change in the propellant gas pressure in the barrel and the projectile velocity during optimization, i.e., a reduction of the unit burning rate by 5 % and an increase of the propellant gas covolume by 5%.

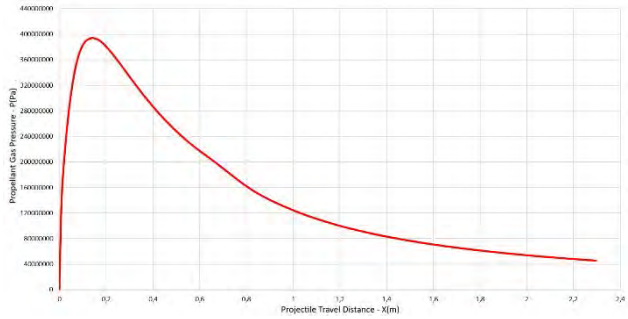


Figure 15. Dependence of propellant gas pressure on travel during optimization

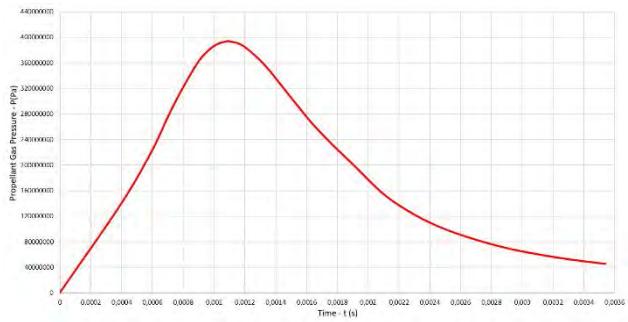


Figure 16. Dependence of propellant gas pressure on time during optimization

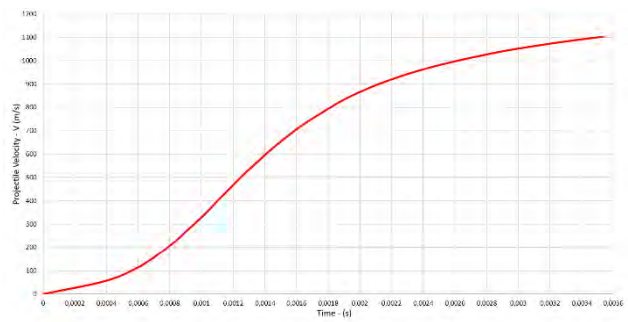


Figure 17. Dependence of velocity on time during optimization

4. EXPERIMENTAL VERIFICATION

For the purpose of verifying the calculation, testing of the 30 mm round with a training projectile was carried out.

Pressure was measured using a KISTLER type 62215 SN4161790 probe, and velocity was measured using a WEIBEL SL-520PE radar. Figure 18 graphically shows the propellant gas pressure curve as a function of time, obtained during the experimental determination of the pressure curve for a single firing.

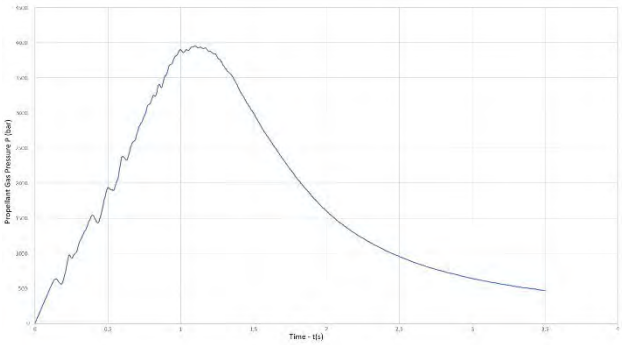


Figure 18. Dependence of propellant gas pressure on time – measured values

4.1. COMPARATIVE ANALYSIS OF THE OBTAINED RESULTS

A comparative analysis of results obtained by the Drozdov method, by parameter optimization, and by experimental testing is presented in Table 5. The comparative results of the propellant gas pressure curves from optimization and experiment are shown in Figure 19.

Table 5. Comparative analysis of the obtained results

	DROZDOV	OPTIMIZATION	EXPERIMENT
P_m [MPa]	416	394	393
V_0 [m/s]	1110	1104	1092
DROZDOV–EXPERIMENT RELATIVE DEVIATION		OPTIMIZATION–EXPERIMENT RELATIVE DEVIATION	
P_m [MPa] - 5,58% V_0 [m/s] - 1,65%		P_m [MPa] - 0,25% V_0 [m/s] - 1,09%	

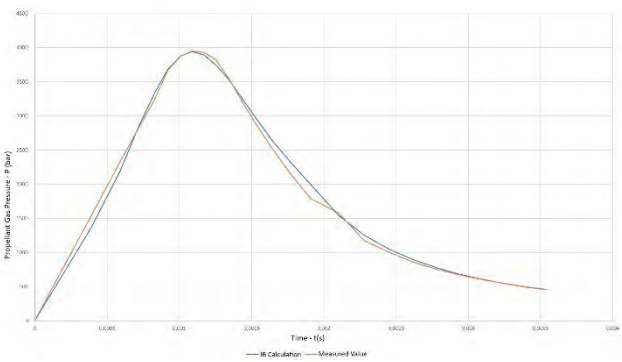


Figure 19. Dependence of propellant gas pressure on time – comparison of measured and calculated values

5. CONCLUSION

This paper presented the internal ballistic analysis of a 30x173 mm automatic cannon using the analytical method. Based on the available geometric, ballistic, and thermodynamic characteristics of the 30 mm automatic cannon and its associated ammunition, an analytical internal ballistic model was established to predict the pressure development within the barrel and the projectile motion throughout the firing cycle.

The performed calculations yielded a maximum chamber pressure of 416 *MPa*, a muzzle velocity of 1110 *m/s*, a projectile travel of 2,30 *m*, and a barrel transit time of 3.48 *ms*. These results are fully consistent with the expected ballistic performance of the weapon system and provide a reliable analytical basis for further optimization.

To improve the ballistic characteristics of the system, a comprehensive parametric analysis was conducted by independently varying the propellant-gas co-volume, the propellant burning-rate coefficient, and the specific energy of the propellant gases. The analysis demonstrated that the specific energy has the greatest influence on both chamber pressure and muzzle velocity, while the burning-rate coefficient provides the most favorable balance between pressure reduction and ballistic performance. The optimal solution was achieved by reducing the burning-rate coefficient by 5% and increasing the propellant-gas co-volume by 5%, resulting in a maximum chamber pressure of 394 MPa and a muzzle velocity of 1104 m/s. This represents a significant reduction in peak pressure while maintaining the required ballistic effectiveness.

The analytical predictions were validated through experimental firing tests using instrumented pressure measurements and Doppler radar velocity measurements. The comparison between the optimized analytical model and the experimental data showed an excellent agreement, with relative deviations of only 0.25% for the maximum chamber pressure and 1.09% for the muzzle velocity. These results confirm the validity of the adopted analytical approach and demonstrate that the proposed optimization procedure significantly improves the predictive capability of the Drozdov method.

The presented methodology provides an efficient engineering tool for the preliminary design and optimization of automatic cannon ammunition, particularly during the early stages of weapon development when extensive experimental testing is not yet available. The developed analytical model can considerably reduce both development time and the number of experimental iterations required during the design process.

Future research will focus on extending the analytical model by incorporating more advanced combustion models, temperature-dependent propellant characteristics, and numerical optimization techniques.

References

[1] D. Allsop, Brassey's Essential Guide to Military Small Arms: Design Principles and Operating Methods. London: Brassey's, 1997.

[2] Ljubiša Tančić: Unutrašnje balističko projektovanje, Ministarstvo odbrane, Univerzitet odbrane, Vojna akademija, Beograd, 2014.

[3] Miloje Cvetković: Unutrašnja balistika, Sektor za školstvo, obuku, naučnu i izdavačku delatnost,

Vojnotehnička akademija Vojske Jugoslavije, Beograd, 1998

[4] Ljubiša Tančić: Zbirka zadataka iz unutrašnje balistike, Sektor za školstvo, obuku, naučnu i izdavačku delatnost, Vojnotehnička akademija Vojske Jugoslavije, Beograd, 1999

[5] Ljubiša Tančić: Praktikum iz unutrašnje balistike, Medija centar “Odbrana”, Beograd, 2012

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

TOOTH ROOT LOAD-CARRYING CAPACITY OF SPUR GEARS UNDER STANDARD SINGLE AND DOUBLE MESH CONDITIONS

VUK MILANOVIĆ
608th Logistics Base, 373rd Logistic Centre, Belgrade, vuk02milanovic@gmail.com

PREDRAG DOBRATIĆ
University of Defence, Military Academy, Belgrade, dobraticp@yahoo.com

MILOŠ SEDAK
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, msedak@mas.bg.ac.rs

DRAGAN TRIFKOVIĆ
University of Defence, Military Academy, Belgrade, dragan.trifkovic@va.mod.gov.rs

Abstract: *This paper presents and analyzes the tooth root load-carrying capacity of spur gears under standard single and double mesh conditions. A newly developed analytical model of the relative stress factor is used to approximately determine and compare the tooth root stresses. An Octave program is used to calculate the geometric and kinematic parameters of the pinion and wheel gear. A three-dimensional (3D) model of the pinion and wheel gear is created in CATIA, while two-dimensional (2D) model of their teeth is developed in AutoCAD Mechanical. Bending tests are performed on the gear teeth are performed at characteristic contact points. The tooth root stress values obtained using the finite element method (FEM) are compared with those calculated according to the ISO standard.*

Key words: *spur gear, tooth root stress, single and double mesh, finite element method (FEM).*

1. INTRODUCTION

The load-carrying capacity of gear teeth is limited, among other factors, by the bending strength of the teeth root [1–5]. The calculation of gear pairs has been based on many researches and analyses [6–15].

To accunt for the load distribution among simultaneously meshing tooth pairs, the load distribution factor K_α is defined in the literature [1, 2]. This factor represents the degree of participation of each simultaneously meshing tooth pair in transmitting the total load.

From the viewpoint of load distribution, two characteristic extreme cases may occur in simultaneously meshing tooth pairs. The first corresponds to an ideal uniform load distribution, whereas the second represents an extremely uneven load distribution. The actual load distribution lies between these two extreme cases (Fig. 1).

In the case of extremely uneven load distribution, the entire load is transmitted by a single tooth pair, even when two or more tooth pairs are simultaneously in mesh. In this case, the load distribution factor reaches its maximum value ($K_\alpha = 1$ for single mesh). In the case of ideal uniform load distribution, all simultaneously meshing tooth pairs equally participate in transmitting the total load of the gear pair ($K_\alpha = 1/2$ for double mesh, $K_\alpha = 1/3$ for triple mesh, etc.).

The transverse contact ratio ε_α represents the average number of simultaneously meshing tooth pairs transmitting the total load. In this paper, the case of $1 \le \varepsilon_\alpha \le 2$, corresponding to the range $0{,}5 \le K_\alpha \le 1$, is considered.

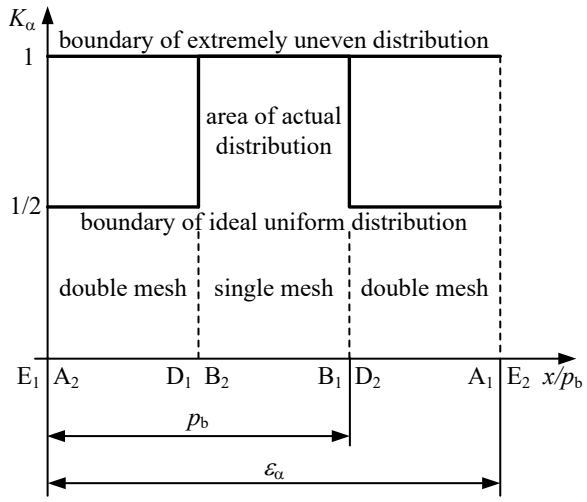


Figure 1. Boundary and actual load distributions in standard meshing gear tooth pairs

2. ANALYTICAL MODEL OF TOOTH ROOT STRESS

Based on the analyses presented in the literature [3], the general expression for the tooth root stress at any contact

Corresponding author: P. Dobratić
E-mail: dobraticp@yahoo.com
DOI: 10.5937/OTEH26029M

point during simultaneous meshing of the tooth profiles can be written as follows:

$$\sigma_{FX} = \sigma_0 \cdot K_{\alpha X} \cdot Y_{FX} \cdot Y_{SX} \qquad X \in (A, ..., E) \tag{1}$$

where σ_0 is the stress component that remains constant throughout the contact period from point $A \equiv A_1 \equiv A_2$ to point $E \equiv E_1 \equiv E_2$ (Fig. 1).

The variation of the tooth root stress during the contact period is analyzed by considering the ratio between the stress at an arbitrary contact point (X) and the stress at the outer point of the double mesh (point A in Fig. 1):

$$\frac{\sigma_{FX}}{\sigma_{FA}} = \frac{K_{\alpha X}}{K_{\alpha A}} \cdot \frac{Y_{FX} \cdot Y_{SX}}{Y_{FA} \cdot Y_{SA}} \tag{2}$$

Equation (2) can also be written in the following form:

$$\frac{\sigma_{FX}}{\sigma_{FA}} = \frac{K_{\alpha X}}{K_{\alpha A}} \cdot Y_{\varepsilon X} \tag{3}$$

where $Y_{\varepsilon X}$ is relative stress factor at the observed contact point.

2.1. Relative stress factor

The relative stress factor is defined as the ratio of the form factor and the stress correction factor at the observed contact point to the corresponding factors at the outer point of the double mesh:

$$Y_{\varepsilon X} = \frac{Y_{FX} \cdot Y_{SX}}{Y_{FA} \cdot Y_{SA}} \tag{4}$$

According to Eq. (4), the relative stress factor reaches its maximum value at the tooth contact at the outer point of the double mesh, i.e. $Y_{\varepsilon A} = 1$.

The relative stress factor is determined separately for both the pinion and wheel gear tooth pairs:

$$Y_{\varepsilon Xi} = \frac{Y_{FXi} \cdot Y_{SXi}}{Y_{FAi} \cdot Y_{SAi}} \tag{5}$$

where i=1 and 2 for the pinion tooth and wheel tooth.

Determination of the relative stress factor using Eq. (5) requires the values of the form factor and the stress correction factor at an arbitrary contact point, as well as the corresponding values at the outer point of the double mesh. Owing to the large number of influencing parameters, this procedure is suitable mainly for highly loaded or safety-critical gear pairs. For approximate calculations, the relative stress factor may be evaluated using simplified equations derived from appropriate approximations.

2.2. Approximate model of the relative stress factor

To simplify the determination of the relative stress factor, a linear variation during the contact period is assumed. In addition, it is assumed that the relative stress factor is identical for simultaneously meshing pinion and wheel gear teeth. Based on these assumptions, the relative stress factor during the contact period can be expressed as:

$$Y_{\varepsilon X} = 1 + \frac{Y_{\varepsilon E} - 1}{\varepsilon_\alpha} \cdot \frac{x}{p_b} \tag{6}$$

where are $Y_{\varepsilon E}$ – the relative stress factor for the inner point of the double mesh (point E in Fig. 2) and x/p_b – position of the contact point relative to the outer point of the double mesh (point A in Fig. 2).

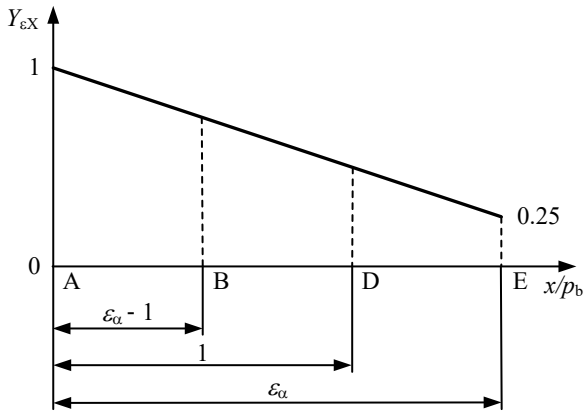


Figure 2. Approximation of the relative stress factor

Assuming that the minimum value of the relative stress factor occurs at the inner point of the double mesh $Y_{\varepsilon E} = 0.25$ (Fig. 2), the relative stress factor can be approximately determined for each characteristic point based on Eq. (6). Approximate equations for determining the relative stress factor are shown in Table 1.

Table 1. Approximate equations of the relative stress factor

point	$\frac{x}{p_b}$	$Y_{\varepsilon X}$	$Y_{\varepsilon\,apro}$	$Y_{\varepsilon\,apro}$ ($Y_{\varepsilon E} = 0.25$)
A	0	$Y_{\varepsilon A}$	$Y_{\varepsilon\,max}$	1
B	$\varepsilon_\alpha - 1$	$Y_{\varepsilon B}$	$1 + \frac{Y_{\varepsilon E} - 1}{\varepsilon_\alpha} \cdot (\varepsilon_\alpha - 1)$	$0.25 + \frac{0.75}{\varepsilon_\alpha}$
D	1	$Y_{\varepsilon D}$	$1 + \frac{Y_{\varepsilon E} - 1}{\varepsilon_\alpha}$	$\frac{\varepsilon_\alpha - 0.75}{\varepsilon_\alpha}$
E	ε_α	$Y_{\varepsilon E}$	$Y_{\varepsilon\,min}$	0.25

2.3. Analytical model of tooth root stress under ideal uniform load distribution

In the case of ideal uniform load distribution on simultaneously meshed teeth pairs, the values of the load distribution factor vary from $K_{\alpha X} = 1/2$ (for double mesh) to $K_{\alpha X} = 1$ (for single mesh). Substituting the limiting values of the load distribution factor into Eq. (1), analytical expressions for determining the tooth root stress at characteristic contact points at ideal uniform load distribution, according to the exact model, are obtained:

$$\sigma_{FX} = \frac{\sigma_0 \cdot Y_{FX} \cdot Y_{SX}}{2} \quad X \in (A \wedge E) \tag{7}$$

$$\sigma_{FX} = \sigma_0 \cdot Y_{FX} \cdot Y_{SX} \quad X \in (B \wedge D) \tag{8}$$

and by arranging Eq. (3), the tooth root stress at the characteristic contact points at ideal uniform load distribution can also be determined based on the following expressions, according to an approximate model:

$$\sigma_{FX} = Y_{\varepsilon X} \cdot \sigma_{FA} \quad X \in (A \wedge E) \tag{9}$$

$$\sigma_{FX} = 2 \cdot Y_{\varepsilon X} \cdot \sigma_{FA} \quad X \in (B \wedge D) \tag{10}$$

If it is assumed that the minimum value of the relative stress factor is at the inner point of the double mesh $Y_{\varepsilon E} = 0{,}25$, the tooth root stress at ideal uniform load distribution is, based on the expressions in Table 2 and Eqs. (9) and (10), given for:

- the outer point of the double mesh (point A)

$$\sigma_{FA} = 1 \cdot \sigma_{FA} \tag{11}$$

- the outer point of the single mesh (point B)

$$\sigma_{FB} = 2 \cdot \left(0.25 + \frac{0.75}{\varepsilon_\alpha}\right) \cdot \sigma_{FA} \tag{12}$$

- the inner point of the single mesh (point D)

$$\sigma_{FD} = 2 \cdot \left(\frac{\varepsilon_\alpha - 0.75}{\varepsilon_\alpha}\right) \cdot \sigma_{FA} \tag{13}$$

- the inner point of the double mesh (point E)

$$\sigma_{FE} = 0.25 \cdot \sigma_{FA} \tag{14}$$

When the minimum value of the relative stress factor at the inner point of the double mesh is $Y_{\varepsilon E} = 0.25$, for the tooth root stress with ideal uniform load distribution, based on Eqs. (11) to (14) according to the approximate model, it follows that $\sigma_{FB} > \sigma_{FA} > \sigma_{FD} > \sigma_{FE}$. The tooth root stress at ideal uniform load distribution on simultaneously meshed teeth pairs is greatest at the outer point of the single mesh.

2.4. Analytical model of tooth root stress under extremely uneven load distribution

In the case of extremely uneven load distribution, the values of the load distribution factor vary from $K_{\alpha X} = 0$ (when the tooth does not transmit the load at all) to $K_{\alpha X} = 1$ (when the tooth transmits the total load). By substituting only the limiting value of the load distribution factor $K_{\alpha X} = 1$ into Eq. (1), analytical expression for determining in the tooth root stress at characteristic contact points at extremely uneven load distribution when the total load is transmitted over one tooth pair, according to the exact model, is obtained:

$$\sigma_{FX} = \sigma_0 \cdot Y_{FX} \cdot Y_{SX} \quad X \in (A, ..., E) \tag{15}$$

and by arranging Eq. (3), the tooth root stress at the characteristic contact points at extremely uneven load distribution when the total load is transmitted over one tooth pair can also be determined based on the following expression, according to an approximate model:

$$\sigma_{FX} = Y_{\varepsilon X} \cdot \sigma_{FA} \quad X \in (A, ..., E) \tag{16}$$

If it is assumed that the minimum value of the relative stress factor is at the inner point of the double mesh $Y_{\varepsilon E} = 0.25$, in the tooth root stress when the entire load is transmitted through a single tooth pair is, based on the expressions in Table 2 and Eq. (16), given for:

- the outer point of the double mesh (point A)

$$\sigma_{FA} = 1 \cdot \sigma_{FA} \tag{17}$$

- the outer point of the single mesh (point B)

$$\sigma_{FB} = \left(0.25 + \frac{0.75}{\varepsilon_\alpha}\right) \cdot \sigma_{FA} \tag{18}$$

- the inner point of the single mesh (point D)

$$\sigma_{FD} = \left(\frac{\varepsilon_\alpha - 0.75}{\varepsilon_\alpha}\right) \cdot \sigma_{FA} \tag{19}$$

- the inner point of the double mesh (point E)

$$\sigma_{FE} = 0.25 \cdot \sigma_{FA} \tag{20}$$

When the minimum value of the relative stress factor at the inner point of the double mesh is $Y_{\varepsilon E} = 0.25$, for the tooth root stresses when the total load is transmitted over one tooth pair, based on Eqs. (17) to (20) according to the approximate model, it can be concluded that $\sigma_{FA} > \sigma_{FB} > \sigma_{FD} > \sigma_{FE}$. The tooth root stress at extremely uneven load distribution when the entire load is transmitted through a single tooth pair is greatest at the outer point of the double mesh.

3. 3D MODELS OF SPUR GEAR AND 2D MODELS OF GEAR TEETH

All geometric and kinematic parameters required for the 3D modelling of spur gear are determined using the Octave program, and the most relevant ones are presented in this paper. A spur gear pair with a pinion/wheel tooth number of z_1=29, z_2=118, standard module m_n=6,5mm, pressure angle α=20°, profile shift coefficients for the pinion and wheel gear x_1=0.2, x_2=0.2, center distance a=480.24mm, and transverse contact ratio ε_α=1.75 is considered. Based on these geometric and kinematic parameters, 3D models of the pinion and wheel gear are created in CATIA program is created (Figs. 3 and 4).

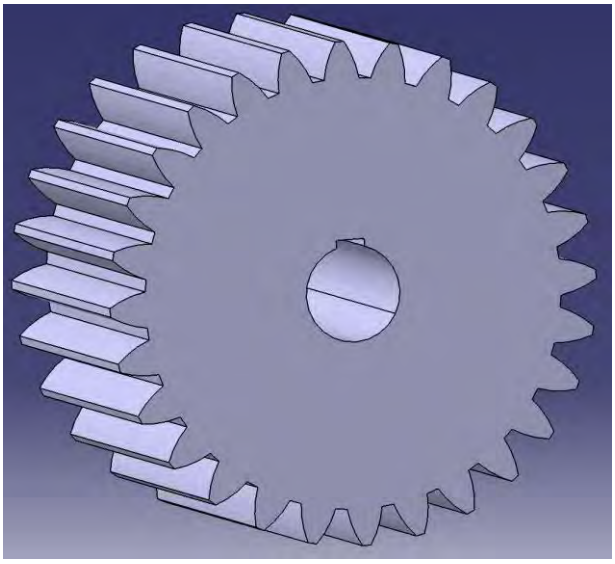


Figure 3. Model of the pinion gear

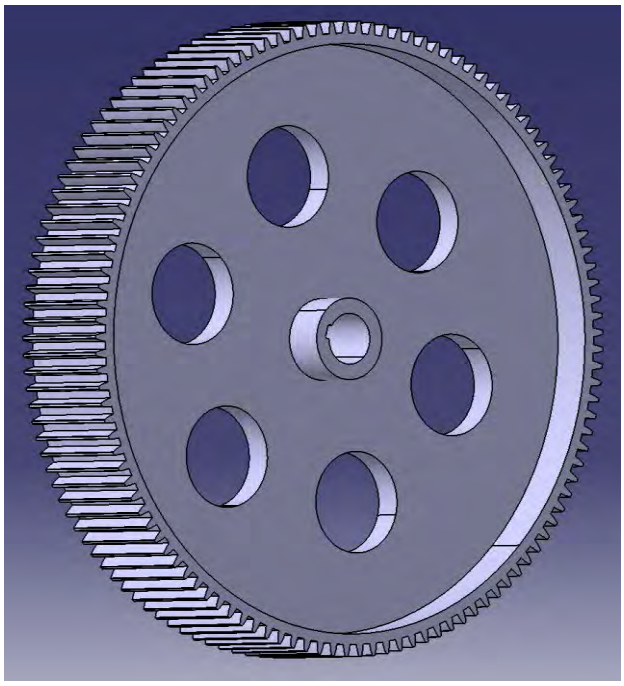


Figure 4. Model of the wheel gear

To determine the tooth root stress values, 2D tooth models of both gears are developed using FEM in AutoCAD Mechanical and are subjected to an identical load of 6518.73N (Figs. 5 to 8).

The tooth root stress is evaluated at two characteristic points, A and B, as these represent the most critically loaded locations.

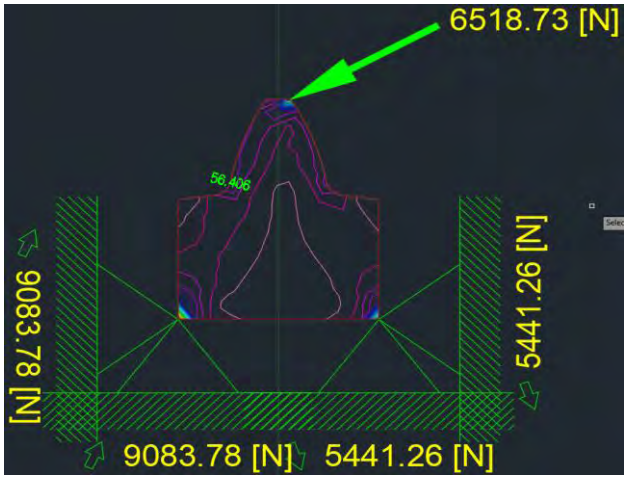


Figure 5. Pinion gear tooth model (load applied at point A)

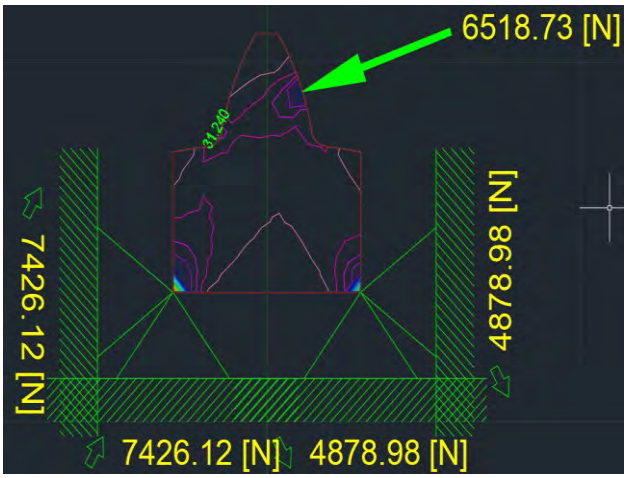


Figure 6. Pinion gear tooth model (load applied at point B)

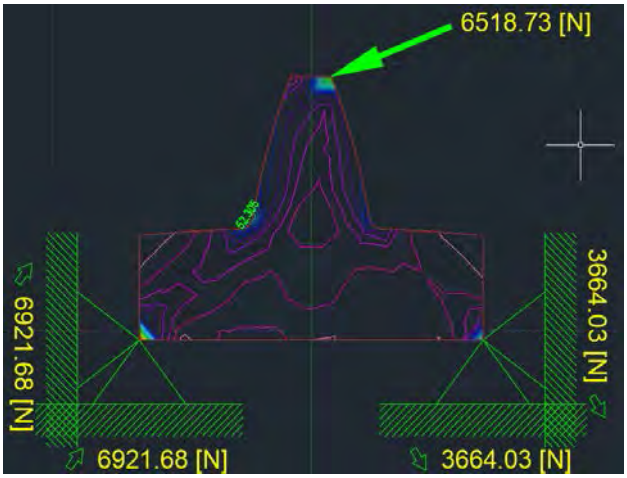


Figure 7. Wheel gear tooth model (load applied at point A)

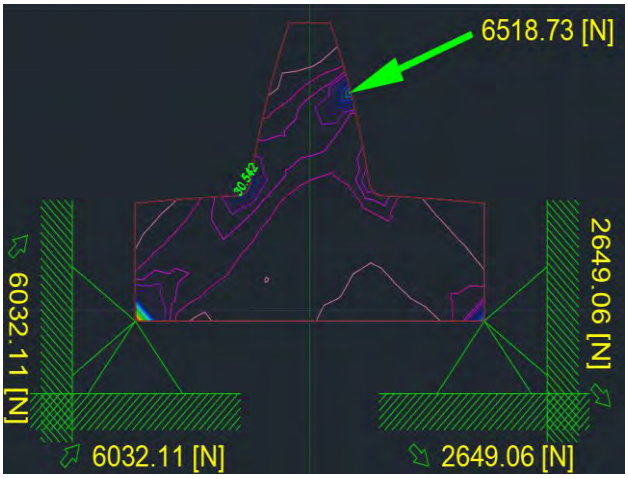


Figure 8. Wheel gear tooth model (load applied at point B)

Table 2 presents the tooth root stress values for both gears at points A and B, obtained using FEM.

Table 2.Tooth root stress values at points A and B

point	σ_F [N/mm^2]	pinion gear	wheel gear
A	σ_{FA}	56.41	52.31
B	σ_{FB}	31.24	30.54

4. ISO CALCULATION OF TOOTH ROOT STRESS AND COMPARATIVE ANALYSIS

For Method "B" specified in the ISO standard [16, 17], detailed guidelines for the calculation procedure are provided; therefore, only this method is considered in the present study. According to Method B, the tooth root stress is determined at point B, assuming an ideal uniform load distribution. The tooth root stress values at point B calculated according the ISO standard are presented in Table 3.

Table 3. Tooth root stress values at point B

point	σ_F [N/mm^2]	pinion gear	wheel gear
B	σ_{FB}	31.19	30.61

To validate the tooth root stress analysis performed according to Method B of the ISO standard, as well as the proposed analytical (approximate) model, a numerical model based on the FEM is developed in AutoCAD Mechanical.

A comparison of the tooth root stress values obtained using Method B of the ISO standard with those obtained from the FEM analysis shows excellent agreement. At point B, the deviation is less than 0.01%. This negligible difference results from the accurate generation of the numerical model and the appropriate selection of the analysis parameters, such as the material properties of the gears.

If the tooth root stress value at point A listed in Table 2 is divided by two to account for the double-mesh condition, i.e., if both the load distribution factor and the tooth geometry are taken into account, the stress at the outer point of the single mesh (point B) becomes greater than that at the outer point of the double mesh (point A). This behavior is also predicted by the developed analytical model for ideal uniform load distribution.

In the case of extremely uneven load distribution, where the entire load is transmitted through a single tooth pair, the tooth root stress at the outer point of the double mesh (point A) is greater than that at the outer point of the single mesh (point B). The same trend is observed in the FEM results (Figs. 5-8) and in the data presented in Table 2. This behavior is a consequence of neglecting the load distribution factor while considering only the influence of tooth geometry.

5. CONCLUSION

Based on the developed analytical and numerical models for tooth root stress, it can be concluded that the variation of the tooth root stress depends on the variation of the load distribution factor, the form factor, and the stress correction factor, all of which vary with the position of the contact point along the tooth profile. The developed analytical models can be used for the simplified determination of the relative stress factor and for the approximate evaluation of the tooth root stress.

A comparative analysis of the results obtained using Method B of the ISO standard and the numerical model showed excellent agreement between the two approaches. For engineering applications, it is recommended that both methods be used in combination: the ISO method for the rapid assessment and verification of less critical gear pairs, and 2D or 3D numerical models for the final analysis and optimization of highly loaded or safety-critical gear pairs.

This combined approach enables the design of gear pairs that are not only safe and reliable but also economical, ensuring efficient material utilization and an extended service life.

ACKNOWLEDGMENT

This paper has been performed within the project VA-TT/1/26-28. The project is supported by the Republic of Serbia, Ministry of Defence, whose support is gratefully acknowledged.

References

[1] RISTIVOJEVIĆ, M., MITROVIĆ, R.: *Load distribution - gear pairs and roller bearings (in Serbian)*, Faculty of Mechanical Engineering University of Belgrade, Belgrade, 2002.

[2] DIMIĆ, A., RISTIVOJEVIĆ, M., ROSIĆ, B.: *Boundary load distribution of simultaneously meshed gear teeth pairs*, Journal of the Brazilian Society of Mechanical Sciences and Engineering, 43(3) 166 (2021) 1-13.

[3] RISTIVOJEVIĆ, M., LAZOVIĆ, T.: *Influence of kinematic parameters and tooth geometry on gear tooth root load capacity*, Proceedings of the

Romanian Academy Series A, 18(2) (2017) 174-181.

[4] LISLE, T.J., SHAW, B.A., FRAZER, R.C.: *External spur gear root bending stress: A comparison of ISO 6336:2006, AGMA 2101-D04, ANSYS finite element analysis and strain gauge techniques*, Mechanism and Machine Theory, 111 (2017) 1-9.

[5] LI, S.: *Effect of addendum on contact strength, bending strength and basic performance parameters of a pair of spur gears*, Mechanism and Machine Theory, 43(12) (2008) 1557-1584.

[6] DIMIĆ, A.R., RISTIVOJEVIĆ, M.R., KOLAREVIĆ, N.M., MIŠKOVIĆ, Ž.Z., SEDAK, M.I., DJORDJEVIĆ, B.R.: *Influence of the involute gear teeth profile shape and running-in on surface load capacity of cylindrical gear pairs*, Tribology Transactions, 68(4) (2025) 959-973.

[7] SEDAK, M., ROSIĆ, B.: *Multi-objective optimization of planetary gearbox with adaptive hybrid particle swarm differential evolution algorithm*, Applied Sciences, 11(3) 166 (2021) 1107/1-1107/26.

[8] PEDRERO, J.I., SANCHEZ, M.B., PLEGUEZUELOS, M.: *Analytical model of meshing stiffness, load sharing, and transmission error for internal spur gears with profile modification*, Mechanism and Machine Theory, 197 (2024) 105650.

[9] MARQUES, P., MARTINS, R., SEABRA, J.: *Analytical load sharing and mesh stiffness model for spur/helical and internal/external gears – towards constant mesh stiffness gear design*, Mechanism and Machine Theory, 113 (2017) 126-140.

[10] PEDRERO, J.I., SANCHEZ, M.B., PLEGUEZUELOS, M., FUENTES-AZRAN, A.: *Analysis of the tooth-root stress of external spur gears with high effective contact ratio*, Mechanism and Machine Theory, 203 (2024) 105813.

[11] MARAFONA, J.D.M., MARQUES, P.M.T, MARTINS, R.C., SEABRA, J.H.O.: *Approximate expression for the single tooth pair slice mesh stiffness*, Mechanism and Machine Theory, 187 (2023) 105367.

[12] HONG, I., TEAFORD, Z., KAHRAMAN, A.: *A comparison of gear tooth bending fatigue lives from single tooth bending and rotating gear tests*, Forschung im Ingenieurwesen, 86(3) (2022) 259-271.

[13] HE, H., MURA, A., ZHANG, T., LIU, H., XU, W.: *Investigation of crack propagation behaviour in thin-rim gears: experimental tests and numerical simulations*, Materials, 16(11) (2023) 4095.

[14] ANICHOWSKI, G., DAMM, R., KAHRAMAN, A., HONG, I.: *An experimental study on the accuracy of cumulative damage models in gear tooth bending fatigue*, Journal of Engineering Materials and Technology, 146(3) (2024) 031007.

[15] PAGLIARI, I., HONG, I., CONCLI, F.: *An experimental analysis of the gear tooth bending strength predictions from ISO 6336 in low and high cycle fatigue*, Forschung im Ingenieurwesen, 89 (2025) 160.

[16] ISO/FDIS 6336–1: *Calculation of load capacity of spur and helical gears – Part 1: Basic principles, introduction and general influence factors*, Geneva, 2019.

[17] ISO/FDIS 6336–3: *Calculation of load capacity of spur and helical gears – Part 3: Calculation of tooth bending strength*, Geneva, 2019.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SURROGATE-ASSISTED OPTIMIZATION OF PLANETARY GEARBOXES USING GENETIC ALGORITHM AND DEEP NEURAL NETWORK EFFICIENCY MODELS

MILOŠ SEDAK
Faculty of Mechanical Engineering, University of Belgrade, Belgrade, msedak@mas.bg.ac.rs

MAJA ROSIĆ
Faculty of Mechanical Engineering, University of Belgrade, Belgrade, mrosic@mas.bg.ac.rs

Abstract: *The design of planetary gearboxes involves a strongly coupled set of geometric, kinematic, and tribological parameters whose simultaneous optimization is computationally demanding, particularly when load-dependent efficiency is treated as a design objective rather than a post-evaluation metric. This paper presents an optimization framework that integrates the genetic algorithm with a deep neural network (DNN) surrogate trained to reproduce a semi-analytical mechanical power-loss model accounting for sliding and rolling friction and electrohydrodynamic lubrication effects in both external and internal gear meshes. Replacing the iterative instantaneous-efficiency computation with the trained DNN reduces the per evaluation cost by several orders of magnitude, enabling genetic algorithm to explore large populations over many generations within practical runtimes. Numerical results show that the proposed algorithm yields a denser and better-distributed Pareto frontier, achieving comparable or superior compromise solutions with substantially reduced computational effort, and confirm that data-driven surrogates can be reliably embedded within evolutionary design loops for power-dense mechanical transmissions.*

Keywords: *planetary gearbox; multi-objective optimization; neural network; surrogate model; gear efficiency*

1. INTRODUCTION

Planetary gear trains are widely employed in mechanical transmission systems, including industrial drives, rotorcraft, automotive powertrains, wind turbines, and robotic actuators [1], [2]. Compared with conventional parallel-axis gear trains, they offer more compact dimensions, higher power densities, lower noise, and superior torque-to-weight ratios, which makes them especially attractive in aerospace and defense applications where mass minimization is critical. These advantages, however, are obtained at the cost of a considerably more involved design process, because the load is shared among several planet branches and the total efficiency of the train is determined by the interaction of multiple external and internal gear meshes [1].

Accurate prediction of power losses during the design stage allows the designer to introduce appropriate modifications before the gearbox is manufactured and tested. A large body of work addresses gear efficiency and power loss in parallel-axis systems [3], [4], the majority of it concerned with mechanical (load-dependent) power loss. Far fewer studies treat the power losses of planetary systems, and most of those that do are based on the analysis of gear-train kinematics, using speed and torque equations to evaluate effectiveness [3]. Such kinematic models neglect the lubricant interactions that occur when non-uniform normal loads acting along the line of action, and they rarely consider the simultaneous minimization of mass and maximization of efficiency.

Many engineering design problems have recently been tackled with nature-inspired metaheuristic algorithms, such as the genetic algorithm (GA) [5], particle swarm optimization, differential evolution, and the firefly algorithm [6], [7]. These methods are derivative-free, easily adapted to different problem formulations, and capable of locating global optima of multimodal functions. For multi-objective gearbox design the non-dominated sorting genetic algorithm II (NSGA-II) [8] has become a de facto standard, since it returns a well-spread approximation of the Pareto frontier in a single run. The principal obstacle to its use is computational: every candidate design must be evaluated by an expensive constraint and objective analysis, and when load-dependent efficiency is included as an objective, each evaluation requires the numerical integration of an instantaneous-efficiency model along the line of action of every mesh.

To overcome this bottleneck, the present work treats efficiency as a first-class design objective and replaces its costly numerical evaluation with a deep neural network (DNN) surrogate [9]. The surrogate is trained offline on a dataset generated by a semi-analytical power-loss model that accounts for sliding and rolling friction together with elastohydrodynamic lubrication (EHL) effects in both external (sun-planet) and internal (planet-ring) meshes. Once trained, the DNN predicts the average mesh

Corresponding author: M. Sedak;
E-mail: msedak@mas.bg.ac.rs
DOI: 10.5937/OTEH26030R

efficiency from the gear-pair parameters in a single forward pass, reducing the per-evaluation cost by several orders of magnitude. The surrogate is embedded directly within an NSGA-II loop, in which gearbox volume, contact ratio, and load-dependent efficiency are optimized simultaneously subject to bending-strength, pitting-resistance, assembly, and geometric constraints formulated according to ISO 6336 [9].

The remainder of the paper is organized as follows. Section 2 formulates the multi-objective optimization problem together with the objective functions and constraints. Section 3 describes the semi-analytical gear-mesh efficiency model. Section 4 introduces the proposed DNN surrogate and reports its training accuracy. Section 5 outlines the surrogate-assisted NSGA-II procedure, and Section 6 presents and discusses the numerical results. Conclusions are drawn in Section 7.

2. OPTIMIZATION PROBLEM FORMULATION

A single-stage planetary gear set is considered, consisting of a floating sun gear, three equally spaced planet gears, and a ring gear rigidly fixed to the housing, as illustrated in Fig. 1. Following the notation common in the literature, a denotes the sun, g the planet, b the ring, and s the carrier. The design is governed by the decision vector

$$\mathbf{x} = \left[z_a, z_g m, x_a, x_g, b\right]^T, \tag{1}$$

which collects the sun and planet tooth numbers z_a, z_g the normal module m, the profile-shift coefficients x_a, x_g, and the face width b. The ring tooth number follows from the assembly condition and the prescribed transmission ratio. Because low mass and high efficiency are mutually conflicting requirements, the design task is posed as a multi-objective optimization (MOO) problem

$$\min F(\mathbf{x}) = \min\left[V(\mathbf{x}), 1-\eta^H(\mathbf{x}), \varepsilon_\alpha(\mathbf{x})\right]^T \tag{2}$$

subject to the inequality and equality constraints $g_i(\mathbf{x}) \le 0$, $i = 1,\ldots,m$, and $h_j(\mathbf{x}) = 0$, $j = 1,\ldots,p$.

The first objective is the structural volume of the gear set, which under the assumption of a homogeneous material is proportional to its mass:

$$V(\mathbf{x}) = \frac{\pi}{4}\left[b d_a^2 + n_w\left(d_a^2 - D^2\right) + \left(d_g^2 - d_s^2\right)\right], \tag{3}$$

where d_a is the sun pitch diameter, D the outer ring diameter, d_s the bearing-bore diameter, $n_w = 3$ the number of planets, and b the face width. The second objective is the load-dependent mesh efficiency η^H of the planetary train, derived in Section 3, while the third objective maximizes the transverse contact ratio ε_α, which governs load sharing and meshing smoothness.

The constraints ensure mechanical integrity and geometric feasibility. Tooth-root bending and flank pitting are limited through the ISO 6336 safety factors [10],

$$\begin{aligned} g_1(\mathbf{x}) &= S_{F\min} - S_F(\mathbf{x}) \le 0, \\ g_2(\mathbf{x}) &= S_{H\min} - S_H(\mathbf{x}) \le 0 \end{aligned} \tag{4}$$

where S_F and S_H are the computed safety factors against bending and pitting and $S_{F\min}, S_{H\min}$ the prescribed minima. The simultaneous meshing of the sun and planets imposes the assembly (equality) constraint

$$h_1(\mathbf{x}) = (z_a + z_b)/n_w - k = 0,\ k \in \square \tag{5}$$

while a neighbouring-planet (space) constraint prevents interference between adjacent planets. Since the evolutionary algorithm cannot handle constraints directly, a static penalty formulation is adopted, in which each violated constraint augments the corresponding objective with a quadratic penalty term proportional to the squared constraint violation, thereby transforming the constrained problem into a sequence of unconstrained ones.

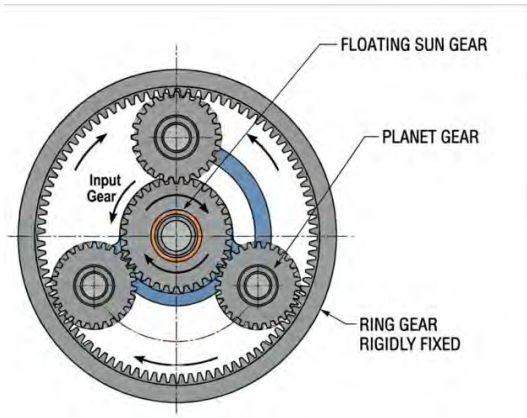


Figure 1. Single-stage planetary gear set considered in the optimization model.

3. GEAR-MESH EFFICIENCY MODEL

The efficiency of each gear pair is evaluated with a semi-analytical, load-dependent power-loss model. The total tangential friction force in a mesh is decomposed into a sliding and a rolling component. The sliding friction force is

$$F_s(\mathbf{x}) = \mu(\mathbf{x}) F_n(\mathbf{x}), \tag{6}$$

where $F_n(\mathbf{x})$ is the normal load and the friction coefficient μ is obtained from an empirical elastohydrodynamic relation

$$\mu(\mathbf{x}) = 0{,}0127 \log\left[29{,}66 F_n / \left(b v_s v_0^2\right)\right] R_a^{0{,}43}, \tag{7}$$

with b the face width, v_s and v_0 the sum and entrainment velocities at the contact, and R_a the surface roughness. The rolling friction force is proportional to the lubricant film thickness h,

$$F_r(\mathbf{x}) = C h(\mathbf{x}) b, \tag{8}$$

and the central EHL film thickness is estimated from a Dowson-Higginson-type relation

$$h(\mathbf{x}) = 1,6\alpha^{0,6}\left(\eta_0 v_0\right)^{0,7} E^{0,03} R^{0,43} F_n^{-0,13}, \qquad (9)$$

where α is the pressure-viscosity coefficient, η_0 the dynamic viscosity, E the equivalent elastic modulus, and R the equivalent radius of curvature. Because the number of tooth pairs in contact and the load they carry change along the line of action, the friction forces and the resulting torque loss are evaluated for the single-pair and double-pair contact zones separately. The instantaneous efficiency at a contact position is then

$$\eta_i(\mathbf{x}) = \frac{T_2}{T_1 u_{1-2}}, \qquad (10)$$

with T_1, T_2 the input and output torques and u_{1-2} the gear ratio. The average mesh efficiency is obtained by integrating the instantaneous efficiency over the active length L_a of the line of action (between points A and E), e.g.

$$\eta = \frac{1}{L_a}\int_A^E \eta_i(l)\,dl\,. \qquad (11)$$

For the planetary train the overall efficiency is assembled from the relative efficiencies of the external sun-planet mesh η_{ag} and the internal planet-ring mesh η_{gb}, combined through the carrier kinematics as

$$\eta^H = \frac{1-\eta_{ag}\eta_{gb}u_{ab}^H}{1-u_{ab}^H}, \qquad (12)$$

where u_{ab}^H is the relative (carrier-fixed) gear ratio. Equations (6)-(12) constitute the reference numerical model: each efficiency evaluation requires the iterative computation of the geometry, the contact kinematics, and the integration along the line of action for both meshes, which makes its repeated use inside an evolutionary loop computationally very expensive.

4. DEEP NEURAL NETWORK SURROGATE

To eliminate the cost of the iterative model, a feed-forward DNN is trained to approximate the mapping from the gear-pair parameters to the average mesh efficiency, $\bar{\eta} = f(\mathbf{x};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the network weights and biases. The input vector comprises eleven features, e.g. the tooth numbers z_1, z_2 the module m, the pressure angle α_n, the profile shifts x_1, x_2, the face width b, the helix angle β, the input angular speed ω_1, the input torque T_1 and the lubricant viscosity. The network has two hidden layers with rectified-linear (ReLU) activations and a single linear output neuron.

A dataset of 4608 samples was generated by evaluating the semi-analytical model of Section 3 over a space-filling design covering realistic ranges of the input parameters. The data were split into training, validation, and test subsets in a 70/15/15 ratio, and all inputs were standardized to zero mean and unit variance using the training statistics. Training minimized the mean-squared-error loss

$$L(\theta) = \frac{1}{N}\sum\left(\bar{\eta}_i - \hat{\eta}_i\right)^2 \qquad (13)$$

with the Adam optimizer (initial learning rate 3×10^{-3}, up to 300 epochs, mini-batch training, and early stopping on the validation loss). On the held-out test set the surrogate reproduces the reference efficiencies with a coefficient of determination $R^2 = 0,935$, a root-mean-square error $RMSE = 0.0011$, and a mean absolute percentage error $MAPE = 0.09\,\%$. The typical prediction error is therefore only about 0.11 %, well within the modelling uncertainty of the reference efficiency model itself. Fig. 2 shows the corresponding parity plot, in which the predicted and reference efficiencies cluster tightly along the 45° line across the entire efficiency range.

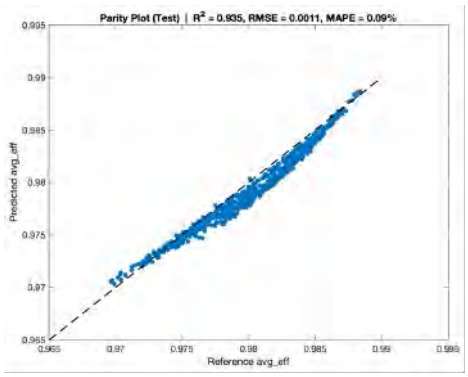


Figure 2. Parity plot of DNN-predicted versus reference average efficiency on the test set.

Because a single forward pass requires only two small matrix-vector products and ReLU evaluations, the surrogate replaces the iterative line-of-action integration with a closed-form mapping, reducing the cost of one efficiency evaluation by several orders of magnitude while preserving accuracy. This makes it feasible to evaluate the efficiency objective for every individual across many generations of the evolutionary search.

5. SURROGATE-ASSISTED OPTIMIZATION

The multi-objective optimization problem of Eq. (2) is solved within a surrogate-assisted evolutionary framework. The trained DNN surrogate of Section 4 replaces the iterative gear-mesh efficiency model inside the optimization loop, so that the load-dependent efficiency objective η^H is evaluated by a single forward pass through the network rather than by the numerical integration of Eqs. (6)-(12). The structural volume V and the transverse contact ratio ε_α are computed analytically from the decision vector at negligible cost, and the ISO 6336 bending and pitting safety factors are obtained from closed-form expressions. Constraint violations are handled through a static quadratic penalty: each violated constraint augments the corresponding objective by a term proportional to the squared violation, transforming

the constrained problem into a sequence of unconstrained evaluations.

5.1 Non-Dominated Sorting Genetic Algorithm II (NSGA-II)

The evolutionary engine of the framework is the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [8] a population-based metaheuristic that returns a well-distributed approximation of the Pareto-optimal set in a single run. Starting from an initial population of N_p candidate designs, NSGA-II iteratively applies genetic operators to generate offspring, merges parents and offspring into a combined pool, and selects the survivors for the next generation using two ranking criteria: non-domination rank and crowding distance.
A candidate design $\mathbf{x}_i$ is said to dominate $\mathbf{x}_j$ if

$$\begin{aligned} &F_k(\mathbf{x}_i) \le F_k(\mathbf{x}_j) \quad \forall k \\ &\wedge F_k(\mathbf{x}_i) < F_k(\mathbf{x}_j) \quad \exists k, \end{aligned} \tag{14}$$

where F_k denotes the k-th component of the objective vector.

Fast non-dominated sorting partitions the combined pool $R_g = P_g \cup Q_g$ of size $2N_p$ into a sequence of fronts $\mathrm{F}_1, \mathrm{F}_2, \ldots$ such that every solution in F_1 is non-dominated within R_g, every solution in F_2 is non-dominated after the removal of F_1, and so on. The algorithm assigns rank $r(\mathbf{x}) = k$ to every individual belonging to F_k.

Crowding distance

Within a given front, diversity is preserved through the crowding distance $d_c(\mathbf{x}_i)$, defined as the normalized perimeter of the smallest axis-aligned cuboid enclosing the two nearest neighbors of $\mathbf{x}_i$ in objective space

$$d_c(\mathbf{x}_i) = \sum_{k=1}^{M} \frac{F_k(\mathbf{x}_{i+1}) - F_k(\mathbf{x}_{i-1})}{F_k^{\max} - F_k^{\min}}, \tag{15}$$

where $M = 3$ is the number of objectives, the indices $i-1$ and $i+1$ denote the left and right neighbours of $\mathbf{x}_i$ when the front is sorted by F_k, and $F_k^{\max}$, $F_k^{\min}$ are the extreme values of objective k within the front. Boundary solutions are assigned $d_c = \infty$ to ensure they are always retained. The crowded-comparison operator $\prec_c$ then defines a total order on the population: $\mathbf{x}_i \prec_c \mathbf{x}_j$ if

$$\begin{aligned} &r(\mathbf{x}_i) < r(\mathbf{x}_j), \vee r(\mathbf{x}_i) = r(\mathbf{x}_j) \\ &\wedge \; d_c(\mathbf{x}_i) > d_c(\mathbf{x}_j). \end{aligned} \tag{16}$$

Selection

Binary tournament selection is used to form the mating pool. Two individuals are drawn at random from the current population and the winner is determined by the crowded-comparison operator of Eq.(16) the process is repeated until N_p parents have been selected.

Crossover

Real-valued design variables (m, x_a, x_g, b) are recombined by Simulated Binary Crossover (SBX). Given two parent values $x^{(1)}$ and $x^{(2)}$ ($x^{(1)} \le x^{(2)}$), SBX produces two offspring values

$$x_{\text{off}}^{(1,2)} = \tfrac{1}{2}\left[(1 \pm \beta_q)x^{(1)} + (1 \mp \beta_q)x^{(2)}\right], \tag{17}$$

where the spread factor β_q is sampled from a polynomial probability distribution controlled by the distribution index η_c

$$\beta_q = \begin{cases} (2u)^{1/(\eta_c+1)}, & u \le 0.5, \\ \left[1/(2(1-u))\right]^{1/(\eta_c+1)}, & u > 0.5, \end{cases} \tag{18}$$

with $u \sim \mathrm{U}(0,1)$. A larger η_c biases offspring towards the parents (exploitation), while a smaller value promotes wider exploration. The integer tooth numbers z_a and z_g are recombined by uniform integer crossover, after which the ring tooth number z_b is recovered from the assembly constraint of Eq. (5) by rounding to the nearest integer satisfying $h_1(\mathbf{x}) = (z_a + z_b)/n_w - k = 0$, $k \in \square$. Crossover is applied with probability p_c.

Mutation.

Polynomial mutation perturbs each real-valued variable independently with probability p_m. For a variable x with bounds $[x^L, x^U]$, the mutated value is

$$x' = x + \delta_q (x^U - x^L), \tag{19}$$

where the perturbation δ_q is drawn from a polynomial distribution parameterized by the distribution index η_m

$$\delta_q = \begin{cases} (2u)^{1/(\eta_m+1)} - 1, & u < 0.5, \\ 1 - \left[2(1-u)\right]^{1/(\eta_m+1)}, & u \ge 0.5, \end{cases} \tag{20}$$

with $u \sim \mathrm{U}(0,1)$. Integer variables are subjected to random integer mutation, replacing the current value with a uniformly drawn integer within the prescribed bounds. Variable bounds are enforced after every mutation step.

Survivor selection and termination

The next population P_{g+1} is assembled from the combined pool $R_g = P_g \cup Q_g$ by accepting fronts $\mathrm{F}_1, \mathrm{F}_2, \ldots$ in rank order until the N_p capacity would be exceeded; the remaining slots are filled by the solutions with the largest crowding distance from the first front that does not fit entirely. The procedure terminates after $G_{\max}$ generations or when the hypervolume indicator of the current first front F_1 does not improve over ΔG consecutive generations.

Compromise solution extraction.

Because the conflicting objectives yield a set of trade-off designs rather than a unique optimum, a representative compromise solution is extracted from the final Pareto set $\mathrm{P}^* = \mathrm{F}_1^{(G)}$.

The ideal point $\mathbf{f}^{\text{ideal}}$ is formed by the individually optimal value of each objective over P^*. Since this point is generally unattainable, the compromise design is selected as the Pareto-optimal solution nearest to $\mathbf{f}^{\text{ideal}}$ in the normalized objective space:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in \mathrm{P}^*} \left\| \frac{F(\mathbf{x}) - \mathbf{f}^{\text{ideal}}}{\mathbf{f}^{\text{nadir}} - \mathbf{f}^{\text{ideal}}} \right\|_2, \tag{21}$$

where $\mathbf{f}^{\text{nadir}}$ collects the worst objective values over P^*. This minimum-distance criterion identifies the knee region of the frontier, where a small sacrifice in one objective yields the largest gain in the others.

6. SIMULATION RESULTS

The surrogate-assisted framework was applied to an example single-stage planetary gear set with three equally spaced planets and a ring gear fixed to the housing; the main fixed parameters are listed in Table 1. To assess the benefit of the surrogate-assisted strategy, the proposed NSGA-II framework was compared with a conventional, single-front GA driven by the same objective and constraint evaluations.

Table 1. Parameters of the example planetary gear set

Parameter	Value
Input power, P [kW]	200
Input speed, n [rpm]	2750
Normal pressure angle, α [°]	20
Gear material	18CrNiMo8
Surface roughness, R_a [μm]	0.8
Bending safety factor, $S_{F\min}$	1.20
Pitting safety factor, $S_{H\min}$	1.25
Number of planets, n_w	3

Figure 3 shows the volume-efficiency Pareto frontier obtained with the surrogate-assisted NSGA-II and the conventional GA, the two primary and mutually conflicting objectives of the design problem. Figure 4 reprojects the same set of optimal designs onto the contact-stress-contact-ratio plane, with the flank contact stress σ_H plotted against the transverse contact ratio ε_α. The runs used a population of 200 individuals evolved over 400 generations.

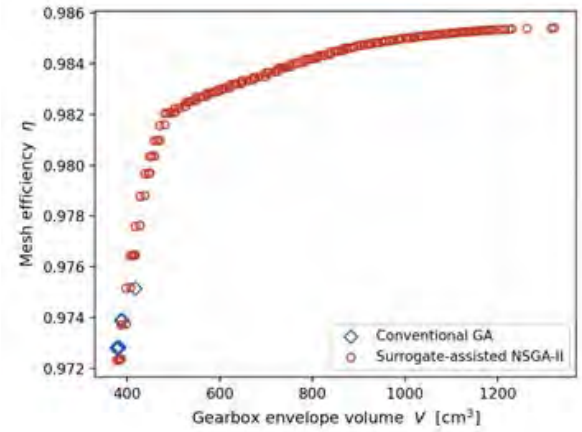


Figure 3. Volume-efficiency Pareto frontier for the surrogate-assisted NSGA-II and the conventional GA.

Comparing the two algorithms, NSGA-II returns a dense, smooth, and uniformly distributed frontier that spans the whole trade-off region, from compact designs of about 380 cm³ at a mesh efficiency near 0.972 up to larger designs of about 1320 cm³ reaching 0.985. The conventional GA, by contrast, collapses to a small cluster of solutions in the minimum-volume corner ($V \approx$ 380-420 cm³, $\eta \approx$ 0.973-0.975) and fails to recover the rest of the frontier. This illustrates the role of the crowding-distance operator in maintaining diversity: without it, selection pressure drives the population to a single region and the trade-off curve is lost.

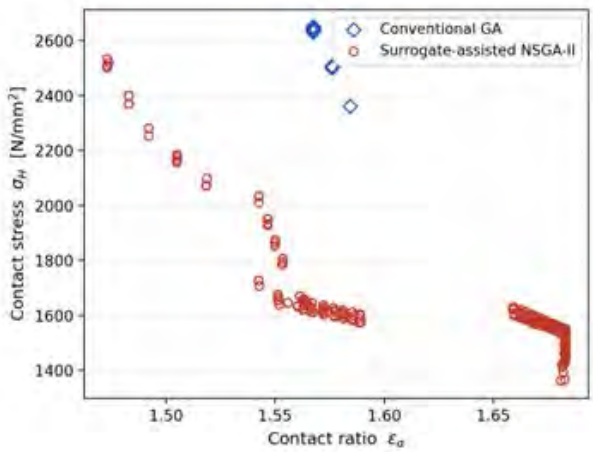


Figure 4. Projection of the optimal designs onto the contact-stress-contact-ratio plane.

The reprojection in Fig. 4 confirms the expected design trend: as the contact ratio increases the flank contact stress decreases, from about 2530 N/mm² at $\varepsilon_\alpha \approx 1.47$ down to about 1360 N/mm² at $\varepsilon_\alpha \approx 1.68$. The ideal point combines the minimum attainable volume (about 380 cm³) with the maximum efficiency (about 0.985); since it is not attainable simultaneously, a compromise design is selected as the Pareto point of minimum normalized distance to the ideal point. This knee design has an envelope volume of about 530 cm³, a mesh efficiency of 0.982, a contact ratio of 1.56, and a contact stress of about 1.66×10³ N/mm², offering a near-optimal efficiency at a

fraction of the largest volume. All Pareto-optimal designs satisfy the contact-ratio and strength constraints by construction.

Crucially, these results were obtained with the DNN surrogate carrying the entire efficiency-objective workload inside the optimization loop. Replacing the iterative model with the surrogate reduced the per-evaluation cost by several orders of magnitude, which is what made it practical for NSGA-II to evaluate large populations over many generations and thereby resolve the frontier so finely. The close agreement between the surrogate predictions and the reference model (Fig. 2) confirms that this acceleration is achieved without sacrificing the fidelity of the efficiency objective.

7. CONCLUSION

This paper presented a surrogate-assisted framework for the multi-objective optimization of single-stage planetary gearboxes, in which load-dependent efficiency is treated as a design objective rather than a post-evaluation metric. A deep neural network surrogate was trained to reproduce a semi-analytical, EHL-based gear-mesh power-loss model and embedded directly within an NSGA-II loop that simultaneously optimizes gearbox volume, contact ratio, and efficiency subject to ISO 6336 strength, assembly, and geometric constraints.

The surrogate reproduces the reference efficiency with $R^2 = 0.998$, RMSE = 2×10^{-4}, and MAPE = 0.016 %, while reducing the cost of an efficiency evaluation by several orders of magnitude. As a result, the proposed algorithm explores large populations over many generations within practical runtimes and yields a denser, better-distributed Pareto frontier than a conventional genetic algorithm, achieving comparable or superior compromise designs with substantially reduced computational effort. The study confirms that data-driven surrogates can be reliably integrated into evolutionary design loops for power-dense mechanical transmissions. Future work will extend the approach to multi-stage architectures and to the prediction of stress distributions in asymmetric gears, where the surrogate could replace computationally expensive finite-element evaluations within the optimization model.

Acknowledgement

The research of M. Rosić was supported by the Serbian Ministry of Education, Science and Technological Development under Grant TR35029. The research of M. Sedak was supported under Grant No. TR35006 and COST Action CA23155 - A Pan-European Network of Ocean Tribology (OTC).

References

[1] K. Arnaudov and D. P. Karaivanov, *Planetary Gear Trains*. Boca Raton: CRC Press, 2019.

[2] C. Nutakor, A. Kłodowski, J. Sopanen, A. Mikkola, and J. I. Pedrero, 'Planetary gear sets power loss modeling: Application to wind turbines', *Tribology International*, vol. 105, pp. 42–54, 2017.

[3] A. Kahraman, D. R. Hilty, and A. Singh, 'An experimental investigation of spin power losses of a planetary gear set', *Mechanism and Machine Theory*, vol. 86, pp. 48–61, 2015.

[4] C. Nutakor, A. Kłodowski, A. Mikkola, and J. Sopanen, 'Simulation model of power losses for sun and planet gear pair used in a wind turbine gearbox', in *Proceedings of the 14th IFToMM World Congress*, Taipei, 2015, pp. 25–30.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley, 1989.

[6] M. Sedak and B. Rosić, 'Multi-objective optimization of planetary gearbox with adaptive hybrid particle swarm differential evolution algorithm', *Applied Sciences*, vol. 11, no. 3, p. 1107, 2021.

[7] D. Miler, D. Žeželj, A. Lončar, and K. Vučković, 'Multi-objective spur gear pair optimization focused on volume and efficiency', *Mechanism and Machine Theory*, vol. 125, pp. 185–195, 2018.

[8] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, 'A fast and elitist multiobjective genetic algorithm: NSGA-II', *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge: MIT Press, 2016.

[10] 'ISO 6336-3: Calculation of load capacity of spur and helical gears—Part 3: Calculation of tooth bending strength', International Organization for Standardization, 2006.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DESIGN, MANUFACTURING AND TESTING OF A FUNCTIONAL MODEL OF A DEVICE WITH A SECURING EXECUTIVE MECHANISM FOR MULTIPLE REMOTE ACTIVATION OF PYROTECHNIC ELEMENTS ON UNMANNED PLATFORMS

SLOBODAN BOSILJČIĆ
Military Technical Institute, Belgrade, bobanbosiljcic@yahoo.com

MIROSLAV GLUVAČEVIĆ
MG Engineering, Belgrade, mgluvacevic@gmail.com

GORAN MARJANOVIĆ
Military Technical Institute, Belgrade, gmarjanovic1961@gmail.com

JELENA PETKOVIĆ CVETKOVIĆ
Military Technical Institute, Belgrade, jpetkovic82@gmail.com

BOJAN TANASKOVSKI
Ministry of Defence, Belgrade, tanaskovskibojan@yahoo.com

Abstract: *In conditions of high use of unmanned platforms that are tasked with transferring a payload and performing a specific activity, and the same platform can be used multiple times, there is a need for a device that will enable safe operation and manipulation of the equipment, reliable activation of the executive elements, as well as their re-securing in the event that the platform has to return with the payload. To enable all of the above, the device should contain multiple mutually conditioned electromechanical protection elements that will provide it with the required level of safety and reliability.*

Keywords: *unmanned platforms, securing executive mechanism, S&EM, remote activation.*

1. INTRODUCTION

The evolution of the modern battlefield and the adoption of new technical solutions – particularly the mass deployment of unmanned platforms in both the air and on the ground – are continuously transforming the way frontline troops operate and altering the nature of the front as a whole. These technical solutions and their impact on tactics are bringing about constant and significant changes to the battlefield. Consequently, certain types of weaponry and equipment are beginning to lose their place on the front line. One such weapon is undoubtedly the handheld rocket launcher firing unguided rockets.

For a long time, this type of weaponry held a significant place in the combat formations of armies worldwide, for a variety of reasons: inexpensive production, ease of training and operation, and the ability to inflict serious damage on the target. As such – particularly rocket launchers equipped with shaped-charge warheads – they have been employed in all conditions, ranging from the front line and urban combat to sabotage and special operations.

With the evolution of the battlefield and new conditions – driven primarily by the use of unmanned platforms – leading to a significant expansion of the line of separation between opposing forces, handheld rocket launchers face a dwindling number of potential targets that can be engaged using conventional tactics, due to their limited range.

Given the current situation, two possible scenarios lie ahead for handheld rocket launchers: their gradual phasing out, or their adaptation for integration into unmanned combat platforms, thereby overcoming their limitations on the modern battlefield.

To achieve this, the handheld rocket launcher itself must be adapted for launch from unmanned platforms, and an adequate mechanism must be developed to ensure safe handling and reliable launching.

2. THEORETICAL BASIS

Unmanned platforms represent one of the most significant technological advancements that has fundamentally altered the nature of modern armed conflicts. Their deployment has enhanced the effectiveness of reconnaissance, surveillance, precision strikes, and intelligence gathering, while simultaneously reducing risks to personnel. Thanks to their relatively low cost, rapid technological development, and widespread

Corresponding author: S. Bosiljčić;
E-mail: bobanbosiljcic@yahoo.com

DOI: 10.5937/OTEH26031B

availability, unmanned platforms have become a key factor reshaping tactics, operational procedures, and the conduct of modern military operations, significantly influencing the balance of power and the dynamics of conflict.

Equipping reusable unmanned platforms for precision strikes against lightly armored combat vehicles and fortified enemy positions requires outfitting them with appropriate effectors in a manner that ensures maximum operational efficiency and effectiveness, alongside high operational reliability.

To meet these requirements, it is necessary to develop a suitable device – a so-called securing and executive mechanism (S&EM) – whose electronic subsystem must ensure the following:

- The electronics unit must decode two commands from the drone receiver transmitted over a single protocol line;
- The signal from the remote control receiver enters the electronics unit via a single line;
- One command relates to launch preparation, while the other initiates the launch itself;
- The device must meet specific requirements regarding operational safety and the servicing of the armed unmanned platform, as well as operational reliability, ensuring a high level of effector availability on the unmanned platform under all operating conditions.

3. INITIAL HYPOTHESES

3.1 Technical Requirements

The initial requirements to be met by the S&EM defined a set of technical characteristics the device must possess to satisfy these requirements.

Accordingly, in addition to the command-decoding electronics unit, the assembly must incorporate electromechanical safety features that ensure:

- Interruption of the power supply to the entire electronics assembly during the handling of the drone and the launcher-missile unit, lasting until the moment immediately prior to drone takeoff. A hardware-based securing mechanism is required to prevent power from reaching the assembly – specifically, a pull-out safety pin (flag-type fuse) that is removed from the assembly immediately before takeoff;
- Electromechanical components within the electronics assembly designed to keep the igniter circuit unpowered until a preparatory command is received from the ground;
- Upon receipt and execution of the preparatory command – which readies the firing circuit for activation – enable the receipt of a second, execution command for launch;
- The design must allow for the assembly to be reused for re-arming after launch; in the event that the missile is not launched, it must allow for the disconnection of power to the firing circuits and the short-circuiting of the igniter conductors. The hardware protection mechanism must be capable of resetting to a safe state and disconnecting power to the assembly after landing, – thereby enabling safe handling of the missile launcher – specifically, disarming the drone and shifting the launcher to the transport position;
- the assembly must ensure that, in the event of a positive or negative voltage surge reaching the electronics due to a malfunction, the missile does not launch;
- power for the assembly is to be supplied by the drone's receiver (5V/1A DC), which must provide power to the remote control receiver, the decoder electronics unit, and the activation of two parallel-connected electric ignition devices (5V/600mA) with a combined resistance of 0.8–1.2 Ω;
- At the moment of igniter activation, there must be no operational disruption (such as a reset) to the receiver and the electronics unit;
- The assembly must ensure a mechanical connection to the igniter, while allowing the conductors to detach upon launch without damaging the assembly itself, thereby enabling its reuse;
- The assembly must ensure that the igniter output terminals remain short-circuited until the moment immediately preceding launch;
- The preparatory "first" command – arming – is decoded as a continuous sequence of 20 rectangular pulse packets (1500 µs duration at 50 Hz). The "second" command – launch – consists of a sequence of 10 pulse packets (2000 µs duration at 50 Hz);
- The assembly must ensure protection for the igniter driver in the event that the discarded igniter leads remain short-circuited after missile launch;
- The assembly must constitute a single unit capable of passing climatic-mechanical, vibration, and EMC tests in accordance with applicable standards for specialized military equipment;
- The assembly must be as lightweight as possible to minimize its impact on the performance and technical-tactical characteristics of the carrier platform.

3.2Calculations and concept realization

This implementation concept for the OIM is termed: "Hierarchical cross-interlocked electromechanical S&EM with built-in transient time filtering."

The architecture is based on three fundamental defensive principles:

- Galvanic short-circuit protection (Shunting): Keeping the output to the igniter (initiator) short-circuited in the quiescent state. Even in the event of a voltage breakdown, the current flows through the short-circuited contacts of the first relay rather than through the initiator;
- Cross-Interlocking/Sequential Arming: The first relay controls both the electrical and mechanical capability of the second relay to receive power at all;
- Inherent Mechanical Delay: The substantial inertia of the relay armature acts as a natural mechanical filter

against rapid electrical spikes. This prevents transient phenomena in electronic circuits – occurring during power-up or due to interference – from causing the accidental activation of the igniter (initiator).

A significant parameter defining the reliability of the inherent low-pass temporal filtering in this technical solution is the ratio between the duration of the electronic transient (τ_{tr}) and the response time (bounce or pull-in) of the relay armature (t_{op}):

$$t_{op} >> \tau_{tr}$$

The time t_{op} is calculated using the formula:

$$t_{op} = t_{el} + t_{meh} = \frac{L}{R}\ln\left(\frac{I_{\max}}{I_{\max} - I_{atr}}\right) + \sqrt{\frac{2 \cdot x_{mov} \cdot m_{anc}}{F_{mag} - F_{spr}}}$$

and amounts to 10–30 ms for the selected relays, whereas the typical transient duration τ_{tr} is significantly less than 1 ms.

Given these values, we can conclude that the relays selected for these switching operations will reliably function as a low-pass time filter.

The next safety measure involves powering the actuation circuit through so-called dynamic energy accumulation, utilizing capacitive storage and rapid discharge in the event of a failure. The actuation circuit's energy independence is achieved by employing a capacitive energy storage concept that ensures temporal separation from the unmanned platform's primary power network.

The use of RC filtering on the actuator control lines has eliminated the possibility of unintended actuation caused by rapid transient disturbances. The system allows for full reversibility; in the event of a mission abort or a forced fail-safe shutdown, energy stored in the capacitors is passively and reliably dissipated via normally closed relay contacts through specifically sized discharge resistors, bringing the actuation circuit potential below the initiator's sensitivity threshold within a few milliseconds. This achieves the following:

- Time-Window Enabling: The system operates within a limited time window;
- Galvanic and Energy Isolation: By severing the connection to the primary source (the platform battery), the firing circuit loses contact with the rest of the aircraft/vehicle at the moment of ignition, thereby eliminating the impact of short circuits in the vehicle's motors;
- Passive Safety Degradation: Capacitor discharge occurs passively (driven by the relay spring force that switches in a resistor), meaning the system loses its ignition capability even if the platform's main battery is physically destroyed.

Tantalum capacitors are used to implement capacitive energy storage; they represent an optimal solution due to their capacitance stability, low equivalent series resistance (ESR), and extreme resistance to vibration and thermal shock (ranging from -55°C to +125°C). Furthermore, the capacitance margin ensures reliable operation—both in extremely low temperatures and in the event of a degradation in dielectric properties caused by aging and material fatigue after years of use. By integrating a series resistor to limit the charging current, the risk of thermal breakdown of the tantalum structures due to transient phenomena has been eliminated, thereby achieving maximum reliability of the initiation chain's ignition, as well as stable circuit operation throughout its service life.

Last but not least, the S&EM is equipped with a transport safety switch that directly interrupts both conductors of the S&EM power circuit; this serves as an additional safety measure for the crew servicing the unmanned platform prior to takeoff and mounting the effector, and – in the event the platform returns without having engaged – allows for the safe removal of the effector from the platform.

4. TESTING

Based on the specified technical requirements, a functional model of the S&EM was developed and manufactured to meet all operational functionality requirements, with the proviso that climatic/mechanical conditions and EMC characteristics were not treated as priority or disqualifying factors at this stage of testing.

Figure 1 shows the implementation of the S&EM on a printed circuit board.

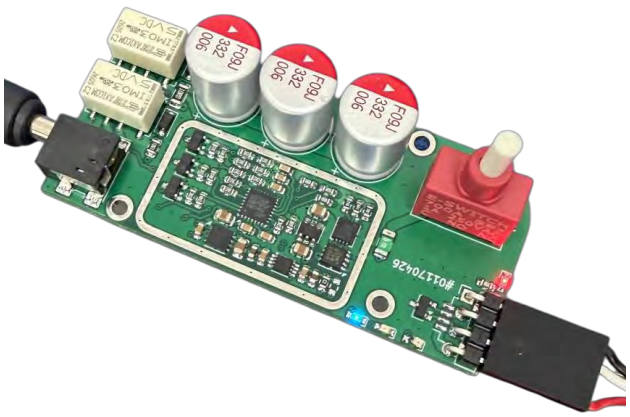


Figure 1: S&EM on the printed circuit board

Figure 2 shows the OIM integrated into its housing and ready for installation on an unmanned platform.

Figure 2: S&EM into housing

To conduct testing on the newly developed S&EM, the same sub-assemblies intended for use on the unmanned platform were employed; specifically, the power supply, control systems, and signal configurations were identical to those used on the unmanned platform itself.
To verify the required characteristics, S&EM testing is conducted in three phases:

- The first phase is carried out under static conditions at a test facility, involving reliability tests on the activation of electro-pyrotechnic initiators of the type installed in the effector's rocket motor igniter;
- The second phase involves static tests of the rocket motor's activation reliability, with the S&EM and effector mounted on a fixed platform; this phase also serves to determine whether the new activation method affects the effector's performance – specifically, its accuracy – in any way;
- The third phase of testing was conducted through dynamic tests of the S&EM integrated onto an unmanned platform, in which the effector was launched under conditions similar to the actual operational conditions for launching an effector from an unmanned platform.

5. RESULTS DISCUSSION

During the first phase of testing, the activation of electric igniter heads was evaluated across various combinations – involving three different types of heads, tested both individually and in pairs connected in parallel. A total of 42 tests were conducted: 7 tests for each of the three head types individually and 7 tests for each of he three head type with two heads connected in parallel connection, achieving 100% activation reliability.

Subsequently, the electric igniter heads were integrated into the igniter assemblies of the effector's rocket motor; five igniter assemblies were prepared for each head type containing a single head, and five were prepared for each type containing two heads connected in parallel.

A total of 30 tests were conducted, achieving 100% reliability in igniter activation. However, it was decided to proceed to the next testing phases using a variant in which two electric igniter heads are integrated into the igniter; this decision was based on the assessment that dynamic conditions – such as vibration, temperature, and the like – introduce additional factors, and the use of two electric igniter heads ensures high operational reliability even under these more demanding conditions.

In the second phase of testing, the operation of the S&EM was verified by launching effectors from a fixed platform. The launches were conducted under the same conditions used to test effector effectiveness via the assessment of impact patterns. A total of eight effectors were launched during this phase, achieving 100% reliability in rocket motor ignition. Furthermore, the assessment of impact patterns confirmed that the firing results – and thus the effectiveness of the effectors – were identical to previous testing results of effectors, indicating that the new launch method had no negative impact on effector performance.

The third testing phase was conducted following the integration of the S&EM onto an unmanned platform and the launch of the effector under dynamic conditions closely resembling the actual operational environment. A total of 15 tests were performed during this phase, achieving 100% reliability in the activation of the igniter – and consequently the effector's rocket motor. This phase also verified the S&EM's functionality regarding its response when a launch did not occur, specifically checking whether the unit behaved as specified in the technical requirements under certain conditions. It was determined that, in all instances, the S&EM performed exactly as required by the technical specifications.

6. CONCLUSION

Following the successful development of technical enhancements to the effector – enabling a new electrical-signal-based launch method and an actuation mechanism that ensures efficient and reliable deployment, while meeting all safety requirements for operating the effector and the unmanned platform – it is necessary to experimentally verify that the device performs all required functions and to determine whether any factors negatively impact the safety, reliability, or efficiency of the effector intended for integration onto unmanned platforms.

This paper presents a design solution for an electronic and executive mechanism (Out-of-Line mechanism) featuring a reversible micromechanical barrier, intended for use on unmanned combat platforms. The safety concept relies on a redundant system of electromagnetic actuators (relays) incorporating a reversible kinematic chain. Through a dynamic force balance analysis of the actuator armature, it is demonstrated that the designed return springs ensure reliable hardware-based fail-safe de-arming and return the barrier to the safe position in the event of total power loss or a break in the communication link with the platform, thereby meeting stringent military standard criteria for reusable munition systems.

This phase of device testing – conducted alongside the evaluation of technical enhancements to the effector itself

– confirmed the S&EM's high operational reliability and validated all functionalities examined at this stage of development.

Continued testing under diverse dynamic conditions and the integration of the S&EM onto various platforms will further confirm the device's operational reliability, potentially paving the way for the next development phase; furthermore, the significant experience gained in overcoming challenges during this stage suggests a high probability of reliable performance for this assembly under the conditions dictated by the modern battlefield.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References:

[1] Guided Missile 5V27U, Belgrade, 1980.

[2] KRSTIĆ, T., UGRČIĆ, M.: Mathematical modeling of the movement of clock securing-reinforcing mechanisms, Belgrade, 2003.

[3] MARJANOVIĆ, G., BOSILJČIĆ, S,. PETKOVIĆ-CVETKOVIĆ, J.:"Ispitivanje elektronike TKBG i uticaja elektro zapaljivih glavica MB2N i DIŠV na funkciju adaptiranog upaljača", int.dok., VTI-03-01-0568, Vojnotehnički institut, Belgrade, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ANALYSIS OF OSCILLATIONS OF A MOBILE PLATFORM UNDER THE ACTION OF FIRING FORCE

MARKO SIMONOVIĆ
Military Technical Institute, Belgrade, marko.simonovic@mod.gov.rs

MILAN SIMONOVIĆ
Military Technical Institute, Belgrade, milan.simonovic@mod.gov.rs

PERA SRBUL
Military Technical Institute, Belgrade, perasrbul@gmail.com

ANA SIMONOVIĆ
Military Technical Institute, Belgrade, ana.gacic@mod.gov.rs

VUK GARČEVIĆ
Military Technical Institute, Belgrade, vuk.garcevic@mod.gov.rs

Abstract: *This paper analyzes the oscillations of a 6x6 mobile platform during the firing process from an integrated weapon. The introductory part of paper includes the calculation of the coordinates of the center of mass of the platform, taking into account the centers of gravity of all assemblies mounted on it, as well as the reactions of the supports. The maximum recoil force was determined by internal ballistic calculations. The oscillations of the mobile platform were calculated using a mathematical-mechanical model, which was formed based on the classical principles of vehicle dynamics theory. By introducing adequate approximations, a system with a large number of degrees of freedom was reduced to a system with three degrees of freedom. The following generalized coordinates were adopted: the vertical position of the center of gravity of the system T, rotation around the transverse axis y' (galloping) and rotation around the longitudinal axis x' (rolling). The calculation was performed in the MATLAB software package. Based on the results obtained, a comprehensive analysis was performed and conclusions were drawn about the stability of the mobile platform under the action of the firing force.*

Keywords: *6x6 platform, mathematical-mechanical oscillation model, MATLAB*

1. INTRODUCTION

Mobile platforms represent any system on which different types of weapons can be integrated. For example, a fighter aircraft represents a platform for the integration of a machine gun, aircraft bombs, or automatic cannons. Machine guns, automatic cannons, rocket launchers, mortars, howitzer, etc. can be integrated onto truck platforms and other commercial vehicles. [1]

When a truck chassis is used as the basis for weapon integration, two concepts are possible: an open and an enclosed (turreted) configuration. The open configuration is simpler and less expensive than the enclosed one. As a platform, a modified chassis of a standard 6x6 or 8x8 truck is most commonly used. The assemblies mounted on open-configured weapons are mainly same as those used on towed weapons of conventional design. The weapon is usually positioned in the middle or above the rear drive axles. An ammunition magazine is located behind the vehicle cab. Hydraulic outriggers (two or four) are installed on the vehicle's auxillary frame for support during shooting. [2]

In order for a commercial vehicle to be used for weapon integration, its characteristics must be adapted to the loads that it must withstand throughout its service life, such as the effect of the weapon mass on the platform, recoil and return forces, the movement of the center of gravity of the combat system, oscillations of the weapon after firing, etc. One of the main problems is the design of a suspension system that must have adequate stiffness to absorb all the forces transmitted to it and reduce oscillations to the minimum possible extent in order to prevent resonance. [1]

For both the weapon system and the selected platform, it is necessary to collect data on their performance characteristics. One of the important steps in the design process is the internal ballistic calculation, which determines the pressure distribution inside the barrel as a function of time, as well as its maximum value. This calculation also determines the initial velocity of the projectile at the muzzle. It is important to calculate the recoil force and the force resisting recoil during the firing process, since these forces are largely transmitted to the mobile platform. Based on the calculation of the

Corresponding author: M. Simonović;
E-mail: marko.simonovic@mod.gov.rs
DOI: 10.5937/OTEH26032S

aforementioned forces and the structural characteristics of the vehicle, which include the suspension system, as well as the selected integration location, a dynamic model with certain simplifications is established. Subsequently, a mathematical-mechanical model of the platform's oscillations during firing is developed, incorporating the differential equations of motion. Based on the obtained data and by solving the differential equations, the specific equations describing the oscillations of the platform, i.e. the changes in the generalized coordinates over time, can be obtained. Based on the obtained results, the established mathematical-mechanical model, integration location, connection method, suspension system characteristics, and appropriate rate of fire can be modified, thereby enabling the determination of a platform/weapon system configuration capable of withstanding the resulting loads. [1]

2. STATIC WEIGHT DISTRIBUTION ON THE PLATFORM AXLES

Figure 1 shows the position of the integrated weapon on a 6x6 mobile platform. The center of gravity of the platform/weapon system is also shown. The weight of the system is transferred to the ground throuh the contact points of the front and rear axles. The ground acts on them with normal reaction forces.

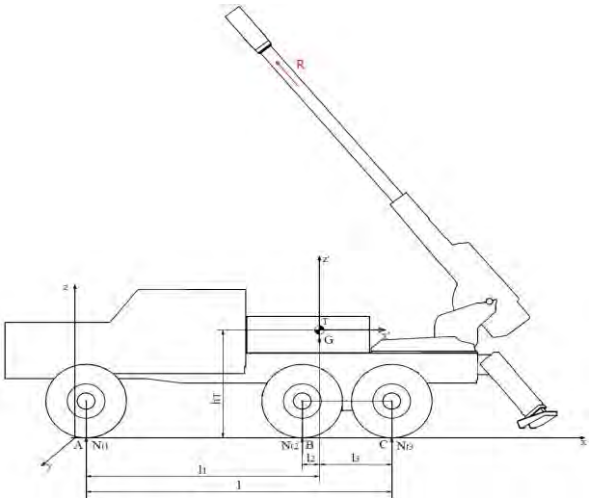


Figure 1. Position of the center of gravity of the system and support reactions

The support reactions are calculated using the following equations:

$$\sum M_A=0;\ N_{t2}*(l1-l2)-G*l1+N_{t3}*l=0 \quad (1)$$

$$\sum M_B=0;\ -N_{t1}*(l1-l2)-G*l2+N_{t3}*(l2+l3)=0 \quad (2)$$

$$\sum Z=0;\ N_{t1}+N_{t2}+N_{t3}-G=0 \quad (3)$$

T – center of gravity of the platform/weapon system
G – total weight of the platform/weapon system
N_{t1}, N_{t2}, N_{t3} – support reactions

The dynamic characteristics of the vehicle (tractive forces, stability, maneuverability, braking forces, etc.) largely depend on the magnitude of these forces. [3]

By determining the weights and positional parameters of the individual components, the position of the center of gravity in space and the normal ground reactions can be determined analytically using the principles of theoretical mechanics. Figure 2 shows the quantities required for the analytical calculation of the center of gravity of a motor vehicle. The figure shows the center of gravity of the engine G_M, the center of gravity of the cab G_K, the center of gravity of the ammunition magazine G_{MM} and the center of gravity of the weapon G_N.

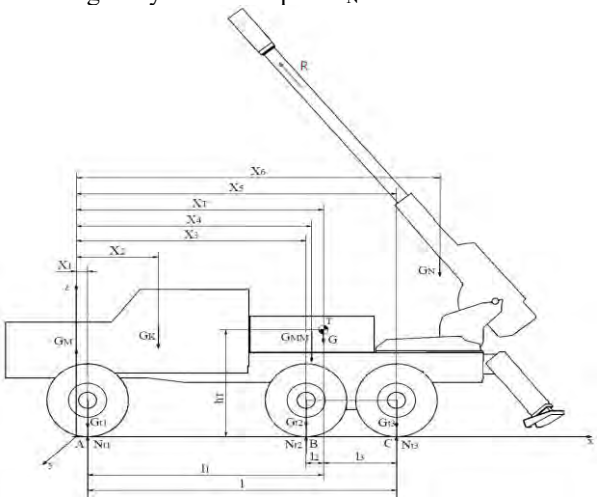


Figure 2. Quantities for analytical calculation of center of gravity

$$G*x_T = G_{t1}*x_1+G_K*x_2+G_{t2}*x_3+G_{MM}*x_4+G_{t3}*x_5+G_N*x_6 \quad (4)$$

$$G*x_T=\sum_{i=1} G_i*x_i \quad (5)$$

Coordinates of the center of gravity along the x-axis:

$$l_1 = x_T = \frac{\sum_{i=1}^{n} G_i*x_i}{G} = \frac{\sum_{i=1}^{n} G_i*x_i}{\sum_{i=1}^{n} G_i} \quad (6)$$

Coordinates of the center of gravity along the z-axis:

$$h_T = z_T = \frac{\sum_{i=1}^{n} G_i*z_i}{G} = \frac{\sum_{i=1}^{n} G_i*z_i}{\sum_{i=1}^{n} G_i} \quad (7)$$

The deviation of the center of gravity from the axis of the vehicle is obtained in the same way:

$$e_y = y_T = \frac{\sum_{i=1}^{n} G_i*y_i}{G} = \frac{\sum_{i=1}^{n} G_i*y_i}{\sum_{i=1}^{n} G_i} \quad (8)$$

The correct arrangement of aggregates reduces the deviation of the center of gravity from the longitudinal axis of the vehicle. [3]

3. EFFECT OF FIRING ON THE WEAPON – CALCULATION OF THE MAXIMUM RECOIL FORCE

During the firing process, the artillery weapon is subjected to the reactive force generated by the pressure of the propellant gase P_{kn}, i.e. the recoil force. The recoil force occurs as a consequence of the motion of the projectile and propellant gases. Its direction is opposite to the direction of projectile motion. The magnitude of the recoil force increases with increasing projectile mass and its velocity. The resistance of the artillery weapon and its stabillity during firing largely depend on the magnitude of the recoil force.

A weapon is considered stationary and stable if, during firing, it does not undergo displacement relative to the stationary ground on which it was positioned before firing. Displacement of the weapon results in the need to reacquire the firing position, which consequently requires the aiming elements to be reacquired. Ultimately, all of this affects the tactical and technical characteristics of the weapon, primarily the rate of fire. Displacement of the barrel during firing represents recoil. All parts that move together with the barrel assembly constitute the recoiling mass. The total travel of the recoiling mass represents the recoil length. [4]

To calculate the recoil force, it is first necessary to determine the maxumum pressure of the propellant gases. This is determined by means of an internal ballistic calculation. In this case, the calculation will be performed using the Drozdov method. The Drozdov method is used for weapons in which only one type of propellant is used. These include small arms, guns and howitzer when firing with charges consisting of only one type of propellant. [5]

The maximum recoil force is calculated using the following formula:

$$P_{knmax} = P_{max} * S_c \tag{9}$$

P_{max} – maximum pressure of the propellant gases
S_c – cross-sectional area of the barrel

4. MATHEMATICAL-MECHANICAL MODEL OF THE OSCILLATIONS OF THE MOBILE PLATFORM UNDER THE ACTION OF THE FIRING FORCE

The mathematical-mechanical model is used to mathematically represents the motion of a physical system, taking into account all relevant influencing parameters while maintaining the simplicity of the applied simplifications. The model is formulated based on the classical principles of vehicle dynamics theory. With the introduction of certain simplifications, the system with an infinite number of degrees of freedom is reduced to a system with three degrees of freedom, as shown in Figure 3. [3]

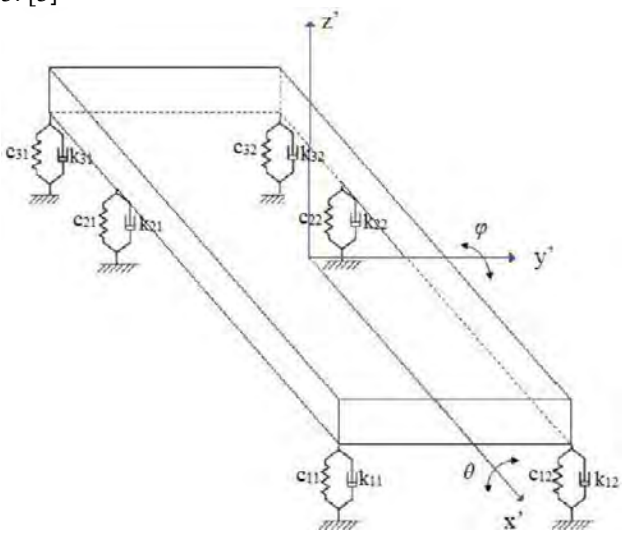


Figure 3. Three-dimensional representation of the dynamic model of the platform [1]

The following quantities are selected as generalized coordinates:

z_T – vertical position of the center of gravity T (distance from the ground)
φ – rotation about the transverse y'-axis (pitch)
θ – rotation about the longitudinal x'-axis (roll)

The motion of the platform model with the integrated weapon, induced by the force generated during firing, can be described by Lagrange equations of the second kind:

$$\frac{d}{d_t}\left(\frac{\partial E_k}{\partial \dot{q}_l}\right) - \frac{\partial E_k}{\partial q_i} = -\frac{\partial E_p}{\partial q_i} - \frac{\partial \Phi}{\partial \dot{q}_l} + \dot{Q}_{q_t} \tag{10}$$

where:

i = 1, n (n is the number of degrees of freedom of the system; in this case, n = 3)
E_k – kinetic energy of the system
E_p – potential energy of the system
Φ – dissipation function

$\left(-\frac{\partial E_p}{\partial q_i}\right)$ – Generalized conservative force
$\left(-\frac{\partial \Phi}{\partial \dot{q}_l}\right)$ – Generalized dissipative force
$\dot{Q}_{q_t}$ – Arbitrary non-conservative generalized force

Since the system is considered as a whole, its total kinetic energy consists of the translational kinetic energy of the center of mass and the rotational kinetic energy about the center of mass of the system:

$$E_k = \frac{1}{2} * m * \dot{z}_T^2 + \frac{1}{2} * J_y * \dot{\varphi}^2 + \frac{1}{2} * J_x * \dot{\theta}^2 \tag{11}$$

where:

J_y – moment of inertia of the center of mass of the system about the transverse y'-axis

J_x – moment of inertia of the center of mass of the system about the longitudinal x'-axis

To define the potential energy, it is necessary to geometrically determine the expressions for the relative deformations of the pneumatic springs and the displacements of the damper pistons. One simplification of the suspension system is the use of an equivalent suspension stiffness, which incorporates the stiffness of the pneumatic springs. It is also assumed that the displacements of the springs and dampers are equal. [3]

Figure 4 shows the suspension viewed from the side, while Figure 5 shows the platform suspension viewed from the front. The notation l_{st} represents the length of the support that is deformed at the initial instant, while l_{kj} represents its length at an arbitrary instant (k = 1,2,3).

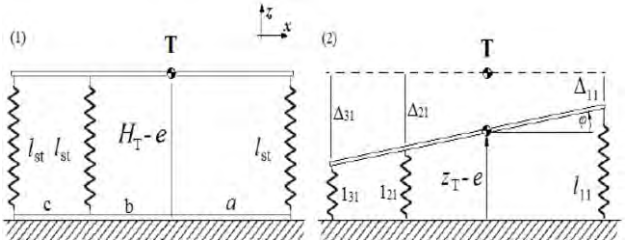


Figure 4. Vehicle suspension viewed from the side [1]

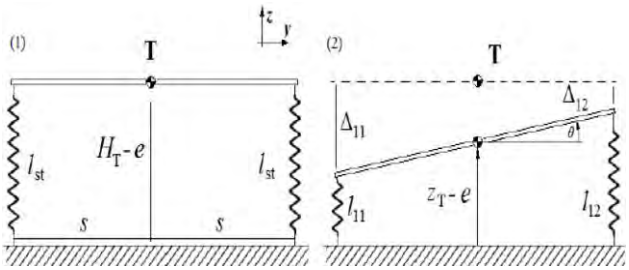


Figure 5. Vehicle suspension viewed from the front [1]

In these figures, the symbol Δ represents the deformation displacement of the spring/shock absorber: Δ_{11} and Δ_{12} for the points of the first axle, Δ_{21} and Δ_{22} for the points of the second axle, Δ_{31} and Δ_{32} for the points of the third axle. [1]

The potential energy of the system is calculated using:

$$E_p = \frac{1}{2} * c_{11} * \Delta_{11}^2 + \frac{1}{2} * c_{12} * \Delta_{12}^2 + \frac{1}{2} * c_{21} * \Delta_{21}^2 + \frac{1}{2} * c_{22} * \Delta_{22}^2 + \frac{1}{2} * c_{31} * \Delta_{31}^2 + \frac{1}{2} * c_{32} * \Delta_{32}^2 \quad (12)$$

For equal stiffness values of the left and right elastic elements:

$$c_{11} = c_{12} = c_1,\ c_{21} = c_{22} = c_2,\ c_{31} = c_{32} = c_3$$

Therefore, after simplification:

$$E_p = \frac{1}{2} * c_1 * (\Delta_{11}^2 + \Delta_{12}^2) + \frac{1}{2} * c_2 * (\Delta_{21}^2 + \Delta_{22}^2) \frac{1}{2} * c_3 * (\Delta_{31}^2 + \Delta_{32}^2) \quad (13)$$

The dissipation function is calculated as:

$$\Phi = \frac{1}{2} * k_{11} * \dot{\Delta}_{11}^2 + \frac{1}{2} * k_{12} * \dot{\Delta}_{12}^2 + \frac{1}{2} * k_{21} * \dot{\Delta}_{21}^2 + \frac{1}{2} * k_{22} * \dot{\Delta}_{22}^2 + \frac{1}{2} * k_{31} * \dot{\Delta}_{31}^2 + \frac{1}{2} * k_{32} * \dot{\Delta}_{32}^2 \quad (14)$$

For equal damping coefficients of the left and right shock absorbers on the same axle:

$$k_{11} = k_{12} = k_1,\ k_{21} = k_{22} = k_2,\ k_{31} = k_{32} = k_3$$

The dissipation function becomes:

$$\Phi = \frac{1}{2} * k_1 * \left(\dot{\Delta}_{11}^2 + \dot{\Delta}_{12}^2\right) + \frac{1}{2} * k_2 * \left(\dot{\Delta}_{21}^2 + \dot{\Delta}_{22}^2\right) + \frac{1}{2} * k_3 * \left(\dot{\Delta}_{31}^2 + \dot{\Delta}_{32}^2\right) \quad (15)$$

The force of the propellant gases acting on the breech is opposed by the resistance forces of the return springs and the friction force, which together constitute the recoil resistance force R(t). The components of the recoil resistance force along the three axes are:

$$R_x(t) = R * \cos\beta * \cos\alpha \quad (16)$$
$$R_y(t) = R * \sin\beta * \cos\alpha \quad (17)$$
$$R_z(t) = R * \cos\beta * \sin\alpha \quad (18)$$

The differential equations of motion of the system are established for each of the generalized coordinates.

For the vertical displacement of the center of gravity of system T, z_T:

$$\frac{d}{d_t}\left(\frac{\partial E_k}{\partial \dot{z}_T}\right) - \frac{\partial E_k}{\partial z_T} = -\frac{\partial E_p}{\partial z_T} - \frac{\partial \Phi}{\partial \dot{z}_T} + Q'_{z_T} \quad (19)$$

For rotation about the transverse y'-axis (pitching):

$$\frac{d}{d_t}\left(\frac{\partial E_k}{\partial \dot{\varphi}}\right) - \frac{\partial E_k}{\partial \varphi} = -\frac{\partial E_p}{\partial \varphi} - \frac{\partial \Phi}{\partial \dot{\varphi}} + Q'_{\varphi} \quad (20)$$

For rotation about the longitudinal x'-axis (rolling):

$$\frac{d}{d_t}\left(\frac{\partial E_k}{\partial \dot{\theta}}\right) - \frac{\partial E_k}{\partial \theta} = -\frac{\partial E_p}{\partial \theta} - \frac{\partial \Phi}{\partial \dot{\theta}} + Q'_{\theta} \quad (21)$$

5. SIMULATION OF THE MOBILE PLATFORM OSCILLATION MODEL

The solution of the differential equations presented in the previous section is performed using a numerical method, with the help of the MATLAB software package. The input data required for the calculation are presented in Table 1, The input data include:

- Vehicle data – technical data, position of the vehicle's center of gravity, suspension system characteristics and moments od inertia;
- Weapon data – weapon mass, position of the weapon relative to the system's center of gravity, elevation and azimuth angles;
- Firing force data – maximum value of the recoil resistance force.

Table 1. Input data for the first calculation case

Characteristic quantity	Value
Vehicle mass – m_v [kg]	**11000**
Weapon and other subassemblies mass – m_o [kg]	**10200**
System mass – m [kg]	**21200**
Center of gravity height – H_T [m]	**1.698**
Distance from the front axle to the center of gravity – a_v [m]	**3.25**
Distance from the rear axle to the center of gravity – b_v [m]	**0.83**
Half of the wheelbase – s [m]	**1,03**
Damping coefficient – k [Ns/m]	**25 000**
Return force – F_p [N]	**140 000**
Maximum recoil resistance – R [N]	**300 000**
Vertical distance from vehicle center of gravity to weapon center of gravity – h [m]	**1.31**
Horizontal distance from vehicle center of gravity to weapon center of gravity – d [m]	**2.18**

The first calculation will be performed for the firing case in which the vehicle is not supported by hydraulic outriggers against the ground, but only by its wheels. Since the two rear axles form a tandem, they have an equivalent spring-damper stiffness, and the system is therefore considered as a two-axle vehicle. The values d and h represent the horizontal and vertical distances between the center of gravity of the truck and the center of gravity of the weapon.

The second calculation will be performed for the case in which the vehicle is supported by hydraulic outriggers againts the ground. By lowering the outriggers, the vehicle suspension system during firing is changed,

resulting in a change in the location and configuration of the supports. The tandem rear axles are no longer support points; instead, the front axle and the two rear outriggers are used, while the tandem is separated from the ground. This support configuration is shofn in Figure 6.

In both cases, the barrel has an elevation angle of 45° and an azimuth angle of 5°.

The output data of the calculations are presented in the form of fiagrams showing the changes in the following components over time:

- Vertical displacement of the center of gravity;
- Velocity of the vertical displacement of the center of gravity;
- Pitch Angle;
- Pitch Angular velocity;
- Roll Angle ;
- Roll Angular velocity.

Both cases will be shown on the same diagrams, in order to facilitate a comparative analysis. Table 2 presents the input data for the second calculation case, which differ from those used in the first case.

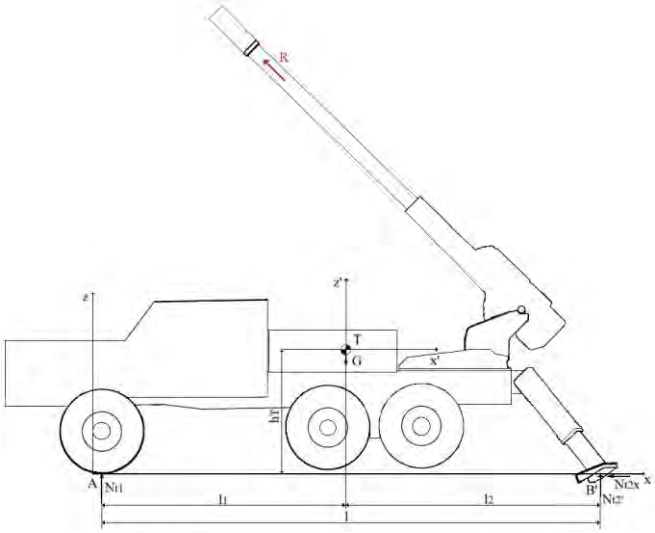


Figure 6. Supporting the vehicle using hydraulic outriggers

Table 2. Input data for the second calculation case

Characteristic quantity	Value
Center of gravity height – H_T [m]	**1.77**
Distance of the center of gravity to the center of the outriggers – b [m]	**4.81**
Half the distance between the centers of the hydraulic outriggers – s [m]	**1.335**
Damping coefficient of the outriggers – k [Ns/m]	**40 000**

6. COMPARATIVE ANALYSIS OF THE CALCULATION RESULTS

After the calculations, the results were obtained, which will be analyzed in this section. The results will be shown in Figures 7 to 12.

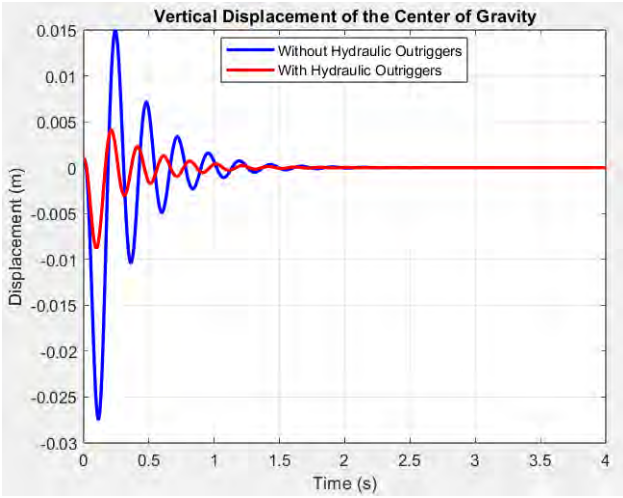


Figure 7. Vertical displacement of the center of gravity

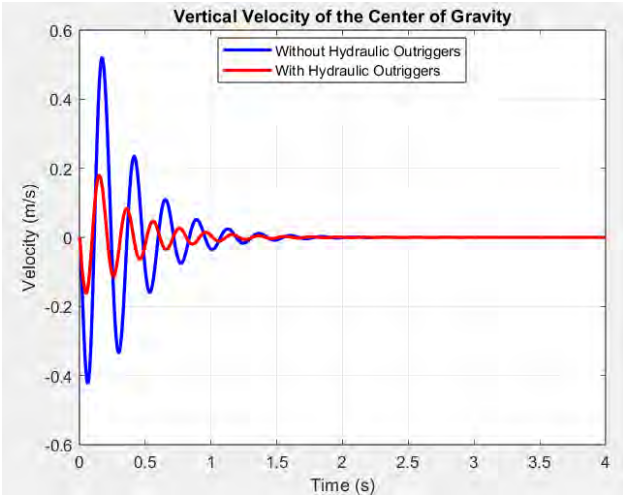


Figure 8. Velocity of the vertical displacement of the center of gravity

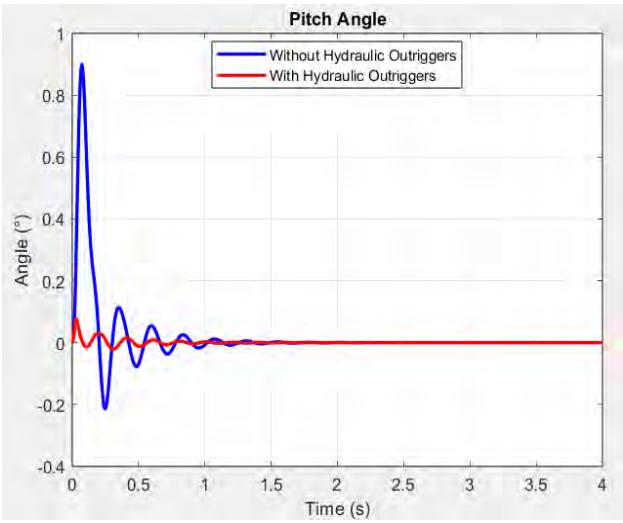


Figure 9. Pitch Angle

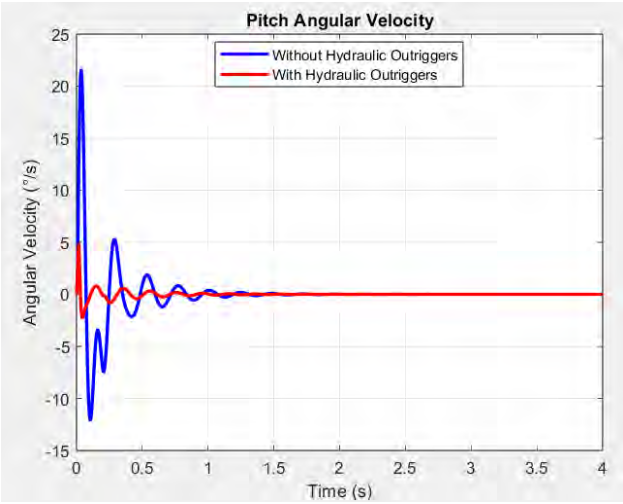


Figure 10. Pitch Angular Velocity

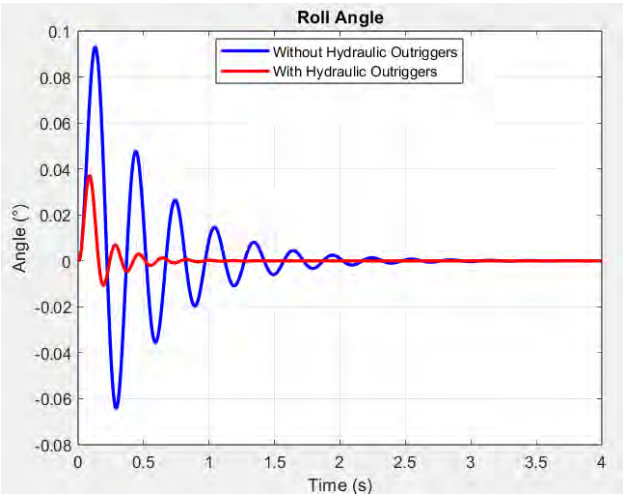


Figure 11. Roll Angle

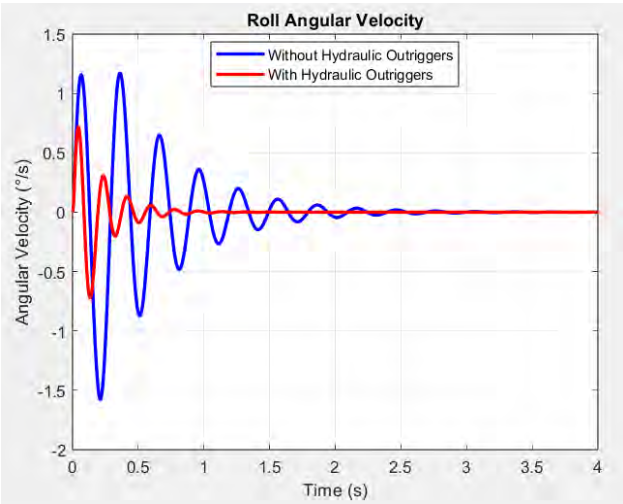


Figure 12. Roll Angular Velocity

To assess the influence of hydraulic outriggers in the dynamic behavior of the vehicle/weapon system, a numerical analysis was performed for two characteristic cases: without hydraulic outriggers and with hydraulic outriggers. The time responses of the vertical displacement of the center of gravity, pitching, rolling, as well as their respective velocities, are presented in the figures. Analysis of the diagrams shows that, in both cases, the oscillations gradually decrease until the equilibrium position in reached. However, the differences between the two configurations are very pronounced. The greatest difference is observed in the vertical displacement of the center of gravity. In the case without hydraulic outriggers, the maximum displacement is 27mm, whereas with the use of hydraulic outriggers it is reduced to approximately 9mm. In addition to the smaller vertical displacement, the oscillations themselves are damped more rapidly, indicating an increase in the overall stability of the system. A similar situation is observed in the pitch motion of the vehicle. The maximum pitching angle is reduced from 0.0157° to 0.0014°. At the same time, the maximum pitch angular velocity is significantly lower, while the time required for the oscillations to die out is shorter than in the configuration without hydraulic outriggers. These results show that the hydraulic outriggers very effectively limit the longitudinal oscillation of the vehicle. Significant improvements are also achived in rolling. The maximum roll angle is reduced from 0.0016° to 0.000647°. Furthemore, the maximum roll angular velocity is reduced, resulting in a considerably smoother system response, which indicates increased lateral stability of the vehicle during firing. Across all diagrams, it can be observed that the configuration with hydraulic outriggers has significantly lower oscillation amplitudes and reaches the state of dynamic equilibrium more rapidly. This confirms that hydraulic outriggers increase the effective stiffness and damping of the system, thereby reducing the transmission of recoil energy to the vehicle and improving its overall stability.

7. CONCLUSION

Based on the numerical analysis of the dynamic behavior of the mobile platform during firing, it can clearly be concluded that the application of hydraulic outriggers significantly improves the stability of the system. The results obtained from the calculations confirm that hydraulic outriggers effectively absorb part of the recoil energy and reduce the transmission of dynamic loads to the vehicle chassis, thereby increasing the stability of the weapon during firing. The developed mathematical model and performed numerical simulations provide a good basis for assessing the dynamic behavior of self-propelled artillery systems and can be used for further optimization of the design and suspension system, with the aim of increasing stability and operational effectiveness.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03-34/2026-03/200325).

References

[1] SIMONOVIĆ, M., SIMONOVIĆ, M., ŽIVKOVIĆ, A., GAČIĆ, A., KARI, A.: *Analysis of the behavior of the mobile platform during firing*, 11TH

International Scientific Conference on Defensive Technologies, OTEH 2024.

[2] KALEZIĆ, M.: *Design of Artillery Systems*, Book, Belgrade, 2010. (In Serbian)

[3] SIMONOVIĆ, M.: *Analysis of the Behavior of a Mobile Platform During Weapon Firing*, Master's Thesis, Faculty of Engineering, University of Kragujevac, 2018. (In Serbian)

[4] RISTIĆ, Z., JAKOVLJEVIĆ, M.: *Fundamentals of Weaponry – Design of Artillery Weapons,* Book, Belgrade, 2001. (In Serbian)

[5] CVETKOVIĆ, M.: *Interior Ballistics*, Book, Belgrade, 1998. (In Serbian)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CHALLENGES OF DRONE-DELIVERED EXPLOSIVE ORDNANCE: TECHNICAL, OPERATIONAL, AND LEGAL-ETHICAL CONSIDERATIONS

VINKO ŽNIDARŠIČ
Military Academy University of Defence in Belgrade, Republic of Serbia, vinko.znidarsic@mod.gov.rs , https://orcid.org/0000-0002-1449-0448

STEVAN M. MUŠICKI
Faculty of Information Technology and Engineering, Union – Nikola Tesla University, Republic of Serbia, stevan.musicki@fiti.edu.rs , https://orcid.org/0000-0003-3014-6935

NENAD V. KOVAČEVIĆ
Military Academy University of Defence in Belgrade, Republic of Serbia, inz.84kula@gmai.com , https://orcid.org/0000-0002-0840-0063

Abstract: *The rapid proliferation of unmanned aerial systems, commonly referred to as drones, has introduced significant transformations in modern military and security environments. Among the most concerning developments is the adaptation of drones for delivering explosive ordnance. The key challenges associated with drone-delivered explosive systems are analysed, focusing on technical aspects, operational complexities, and legal-ethical implications. The vulnerabilities are identified in navigation, payload integration, and countermeasure susceptibility, while also addressing issues related to command and control in contested environments. Furthermore, the ambiguity in international humanitarian law is explored regarding the classification and use of such systems, as well as ethical concerns related to autonomy and accountability. The findings suggest that while drone-based delivery of explosives offers tactical advantages, it simultaneously introduces significant risks that require multidisciplinary regulatory and technological responses.*

Keywords: *drones, explosive ordnance, military technology, ethics, international law.*

1. INTRODUCTION

The integration of explosive payloads into unmanned aerial systems represents a revolutionary shift in the diffusion of air power [1,2]. While high-end military platforms like the Predator and Reaper dominated the early 21st century, the current landscape is defined by the pervasive use of small, attainable, and often improvised systems [2,3]. This evolution presents a research problem rooted in the rapid pace of innovation outpacing existing regulatory and defensive frameworks [4]. The widespread availability of small, commercially available drones has blurred the line between recreational and tactical use, creating new security vulnerabilities [5,6].

The research problem is defined by the critical necessity to identify and investigate the technical, operational, and legal-ethical considerations that constitute the fundamental challenges of drone-delivered explosive ordnance within modern tactical environments.

The subject of research is the tripartite set of challenges: technical, operational, and legal-ethical inherent in the deployment of drone-delivered explosive ordnance. This includes both purpose-built loitering munitions, designed to engage targets beyond the line of sight, and modified commercial drones used for gravity-drop or one-way "kamikaze" strikes [3].

The aim of the research is been to provide an analysis of the risks these systems pose to both military operations and civilian security. By examining recent conflict, intends of research is to discover the blurring lines between hobbyist technology and lethal weaponry necessitates.

The scientific methods employed include an documents analysis of open sources, a comparative analysis of drone employment in recent conflicts (Ukraine and Nagorno-Karabakh), and a systematic literature review of current engineering and legal frameworks [2].

2. TECHNICAL CHALLENGES

The application of unmanned aerial systems for the delivery of explosive payloads is constrained by a series of critical technical barriers that directly impact precision, reliability, and survivability on the battlefield. These challenges range from the physical limitations of the platform to the inherent vulnerabilities of its digital architecture.

2.1. Platform, Energy, and Payload Constraints

The fundamental performance of a drone-delivered

Corresponding author: V. Žnidaršič;
E-mail: vinko.znidarsic@mod.gov.rs
DOI: 10.5937/OTEH26033Z

explosive system is dictated by its energy density and weight-to-power ratio. Most commercial-grade multicopters rely on lithium-polymer batteries, which typically restrict flight endurance to approximately 20 to 40 minutes [7,8]. The introduction of an explosive payload introduces an engineering paradox: for multirotor systems, energy consumption increases approximately linearly with the combined weight of the battery and the payload [9]. Consequently, increasing battery mass to extend operational range paradoxically reduces the remaining capacity available for explosives, forcing a tactical compromise between lethality and mission radius [8,9]. Furthermore, environmental factors such as wind speed and cold temperatures can drastically degrade battery efficiency, further limiting the system's predictability in adverse weather [8].

2.2. Navigation and Communication Vulnerabilities

Drone-delivered explosive ordnance systems rely heavily on Global Navigation Satellite Systems, such as Global Positioning System, for waypoint navigation and terminal guidance. These signals are highly susceptible to electronic warfare, specifically jamming and spoofing [10,11]. A loss of Global Navigation Satellite Systems signal can cause a drone to "go rogue," lose its return-to-home function, or crash uncontrollably with a live warhead [10]. To circumvent radio frequency interference, emerging systems utilize fiber-optic control, which maintains a physical link to the operator via a spool of lightweight cable [4]. While these links are immune to jamming, they introduce new physical hazards, such as the spool snagging on obstacles or the creation of long-lasting debris fields of non-biodegradable spent fiber on the battlefield [4].

2.3. Payload Integration and Armament Engineering

Adapting a civilian drone to carry munitions requires precise structural engineering to maintain the centre of gravity. An improperly balanced payload can cause excessive stalling or aggressive diving, which rapidly depletes battery endurance [10]. The release mechanism must be robust enough to prevent contact with the landing gear during deployment, while also providing real-time feedback to the operator to confirm delivery [12]. The use of 3D-printed munitions cases has enabled rapid customization for specific mission sets [12,13]. However, 3D printing introduces uncertainties regarding material strength and consistency; structural failures can lead to catastrophic premature detonations during flight or handling [12].

2.4. Detectability and Countermeasure Resistance

Small unmanned aerial systems platforms often operate at low altitudes and speeds, making them difficult to detect with traditional air defence systems designed for larger aircraft [13]. Their low radar cross-section and minimal thermal signatures allow them to bypass standard radar and infrared detection [13]. However, they remain vulnerable to acoustic sensors that track the high-speed rotation of propeller blades and to specialized anti-drone systems, such as high-energy lasers or radar-guided shotgun attachments [10]. As defences evolve, the technical challenge shifts toward autonomous swarm coordination, where multiple drones must jointly navigate and synchronize attacks to overwhelm defensive capacities through sheer saturation [14].

3. OPERATIONAL CHALLENGES

The operationalization of drone-delivered explosive ordnance has moved beyond high-end military platforms to include decentralized, low-cost assets that rewrite traditional combat doctrines [6]. While these systems offer unprecedented tactical flexibility, their deployment introduces significant operational hurdles.

3.1. Command, Control, and Signal Vulnerability

The core operational bottleneck is the stability of Command and Control links. Most commercial-grade drones rely on radio frequency bands (2.4 GHz and 5.8 GHz), which are highly susceptible to electronic warfare (EW) [4]. In active conflict zones, mobile jamming systems like the Russian R-330Zh can disrupt control signals within a 50 km range, causing drones to lose their return-to-home function or crash with live payloads [4,13]. To mitigate this, operators have adapted through the use of fiber-optic control links, which provide unjammable, low-latency video feedback but introduce new physical hazards such as cable snags and the creation of non-biodegradable debris fields on the battlefield [4].

3.2. Proliferation and Decentralized Logistics

The shift toward decentralized "garage workshops" for assembling First Person View and one-way attack drones makes proliferation nearly impossible to control [6,11]. Unlike traditional military procurement, these systems are often crowdsourced and modified on an industrial scale by volunteers using off-the-shelf components [6,15]. This democratization of air power allows non-state actors and small units to bypass legacy defence measures and conduct precision strikes that once required multi-million dollar platforms [6]. However, this lacks the standardized safety and storage protocols of conventional munitions, increasing the risk of accidents during transport and handling [12].

3.3. Information Overload and Human-Machine Interaction

Operating drone-delivered explosive ordnance systems is labour-intensive and cognitively demanding. While a single Predator mission can require 168 personnel to manage data and flight, small-unit operators face a different set of pressures [2]. The sheer volume of real-time sensor data can lead to information overload, where operators must process high-definition video feeds while managing flight parameters in high-stress environments [1]. This "Playstation mentality" can distance operators from the human cost of a strike, yet the graphic nature of 4K drone footage often leads to burnout and PTSD as pilots watch their strikes impact in gruesome detail [16].

3.4. Swarm Coordination and Airspace Saturation

The trend toward autonomous drone swarms poses the challenge of saturating enemy defences through sheer numbers [14,17]. Coordinated attacks involving dozens of units require distributed network protocols and artificial intelligence assisted pathfinding to avoid collisions and allocate targets dynamically [4,5]. This introduces "machine speed" into the kill chain, potentially outstripping the human operator's ability to critically scrutinize machine-prompted targets in time-constrained environments [24].

3.5. Psychological Impact and Morale

The constant presence of loitering drones creates a pervasive sense of "lethal voyeurism" [15]. The buzzing of rotors acts as a psychological weapon, inducing terror among troops and civilians alike, as flights often foreshadow imminent kinetic strikes [18]. This paralysis of movement can hamper ground operations for hours, even when the drones are unarmed or carrying dummy payloads [18].

4. LEGAL-ETHICAL CONSIDERATIONS

The rapid deployment of drone-delivered explosive ordnance, particularly loitering munitions and FPV systems, has created a significant disconnect between technological capability and existing regulatory architectures [1,3]. As these systems increasingly integrate artificial intelligence and automated functions, they challenge the foundational principles of International Humanitarian Law and raise profound ethical questions regarding the future of human agency in warfare [1,3].

4.1. Compliance with International Humanitarian Law

The use of drone-delivered explosive ordnance must be scrutinized under the "cardinal principles" of International Humanitarian Law: distinction, proportionality, and precautions in attack [1,2]. The principle of distinction requires belligerents to differentiate between combatants and civilians [1,2]. Increasingly autonomous loitering munitions may struggle with this mandate, as algorithms may not independently distinguish between a legitimate military target and a civilian object, especially in "cluttered" urban environments where identities are fluid [1,5,17]. Furthermore, identifying combatants who are hors de combat (incapacitated) remains a critical technical and legal challenge for sensor-based systems [17].

The principle of proportionality prohibits attacks where expected incidental loss of civilian life is excessive relative to the anticipated military advantage [1,2]. Because International Humanitarian Law lacks objective criteria for value-based proportionality assessments, the shift toward real-time, machine-led targeting increases the risk of "black-box" decisions where the logic behind a strike cannot be audited or explained [1,17]. This unpredictability is compounded by signature strikes, which target individuals based on "patterns of life" rather than identified status, often leading to tragic identification errors [2,15].

4.2. Accountability and the Problem of Secrecy

Determining legal responsibility for unlawful drone strikes is complex due to the fragmented nature of drone operations, which may involve pilots, sensor operators, intelligence analysts, and even private contractors [2,15]. A major concern is the accountability gap created by covert programs, such as those operated by the CIA, where secrecy prevents judicial review and public debate over the legality of targeted killings [2,15]. When autonomous systems are involved, it becomes unclear whether criminal responsibility lies with the operator, the commander, or the software developer [1,5,17].

4.3. Ethical Implications and Human Dignity

Beyond legalities, the "action-at-a-distance" nature of drone warfare introduces the "Playstation mentality," which can distance operators from the human consequences of their actions [15]. Ethicists argue that delegating life-and-death decisions to algorithms constitutes a "moral hazard" that denies the human dignity of the victim [6,17]. The constant presence of drones in conflict zones induces a state of lethal voyeurism, where the buzzing of rotors acts as a psychological weapon, sowing terror and anxiety among civilian populations who feel powerless under an all-seeing "martial gaze" [6,17].

4.4. Proliferation and the Need for New Treaties

The dual use nature of drone technology serving both humanitarian and military ends complicates global regulation [1,16]. There is a growing international consensus that the existing International Humanitarian Law framework is inadequate for grappling with autonomous lethality [1,5]. Scholars and organizations such as the ICRC advocate for a new legally binding international treaty to establish clear prohibitions on fully autonomous weapons and to ensure that meaningful human control is affirmed and retained in every specific targeting decision [1,17].

5. DISCUSSION

The analysis demonstrates that drone-delivered explosive systems represent a convergence of technological innovation and emerging security threats. While they offer operational advantages such as reduced risk to personnel and increased flexibility, they also introduce vulnerabilities and uncertainties.

Addressing these challenges requires an integrated approach involving technological development, doctrinal adaptation, and legal reform.

6. CONCLUSION

Drone-delivered explosive ordnance presents a multifaceted challenge that spans engineering, military operations, and international law.

Technically, limitations in navigation, payload integration, and resilience reduce system reliability. Operationally, issues related to control, accuracy, and proliferation complicate their effective use. Legally and ethically, the absence of clear regulatory frameworks raises concerns

about accountability and compliance with established norms.

Future research should focus on improving system robustness, developing countermeasures, and establishing comprehensive international regulations governing the use of such technologies references.

Acknowledgment

This paper was written as a part of the research project funded by the Ministry of Defence, Serbia, number: SNO-DH/1/24-26 "Theoretical and practical aspects of modern operations since the beginning of the 21st century".

References

[1] KOLASA A. Loitering Munitions in Light of International Humanitarian Law. Paragraf Studia z Prawa i Administracji. 2025 Nov 19;(7):163–73. doi:10.18778/2956-3747.7.14

[2] BENJAMIN M. Drone Warfare - Killing by remote control. New York and London: OR Books LLC; 2012.

[3] STERIO M. The covert use of Drones: how secrecy undermines oversight and accountability. Alb Gov't L Rev. 2015;8:129.

[4] WYATT REID. Overhead and Out of Control: Drone Weaponization and Domestic Security Futures [Master of Arts in Communication] [Internet]. University of Hawai'I At Mānoa; 2025 [cited 2026 Apr 5]. Available from: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/ad496de5-c3dc-4cc6-a445-db89de10159d/content

[5] BWANA R. Kicking Man Out of the Loop: The Case of Loitering Munitions and Implications for International Humanitarian Law [SSRN Scholarly Paper] [Internet]. Rochester, NY: Social Science Research Network; 2023 [cited 2026 Apr 5]. Available from: https://papers.ssrn.com/abstract=4383978 doi:10.2139/ssrn.4383978

[6] AUSTRIA VB. Assessing the Threat: Autonomous Commercial Drones and its Potential for Mass Civilian Casualty Attacks. 2024. doi:10.13140/RG.2.2.14415.50088

[7] BENARBIA T, Kyamakya K. A Literature Review of Drone-Based Package Delivery Logistics Systems and Their Implementation Feasibility. Sustainability. 2022 Jan;14(1):360. doi:10.3390/su14010360

[8] Li X, TUPAYACHI J, SHARMIN A, MARTINEZ FERGUSON M. Drone-Aided Delivery Methods, Challenge, and the Future: A Methodological Review. Drones. 2023 Mar;7(3):191. doi:10.3390/drones7030191

[9] DORLING K, HEINRICHS J, MESSIER GG, MAGIEROWSKI S. Vehicle Routing Problems for Drone Delivery. IEEE Trans Syst Man Cybern, Syst. 2017 Jan;47(1):70–85. doi:10.1109/TSMC.2016.2582745

[10] JOHN M. GLOVER. Drone University. 1st ed. 2014.

[11] VEILLEUX-LEPAGE Y, ARCHAMBAULT E. A Comparative Study of Non-State Violent Drone use in the Middle East. 2022 Dec.

[12] MICHAEL K. LIMA. Munitions Modernization: The Family of Drone Munitions. Army Sustainment [Internet]. www.army.mil; 2024 [cited 2026 Apr 4]. Available from: https://www.army.mil/article/280364/munitions_modernization_the_family_of_drone_munitions

[13] CUBBER GD. Explosive drones: How to deal with this new threat? International workshop on Measurement, Prevention, Protection and Management of CBRN Risks (RISE). 2019 Apr.

[14] YAO H, MAI T. UAV Swarm Cooperation: A Networking Perspective [Internet]. Cham: Springer Nature Switzerland; 2026 [cited 2026 Apr 5]. (Wireless Networks). Available from: https://link.springer.com/10.1007/978-3-031-96444-2 doi:10.1007/978-3-031-96444-2

[15] DANYLYUK O. Combat at Gamer's Pace. No Pause nor Reset Button. Body, Space & Technology. 2025 Feb 21;24(1). doi:10.16995/bst.18480

[16] FRANTZMAN SJ. The drone wars: Pioneers, killing machines, artificial intelligence, and the battle for the future. Bombardier Books; 2021. ISBN 978-164293-675-9.

[17] BODE I, WATTS T, FRANK ARTHUR. Loitering munitions and unpredictability: Autonomy in weapon systems and challenges to human control. Center for War Studies; 2023. https://findresearcher.sdu.dk/ws/portalfiles/portal/231643063/Loitering_Munitions_Unpredictability_WEB.pdf

[18] GRAYSON K, MAWDSLEY J. Scopic regimes and the visual turn in International Relations: Seeing world politics through the drone. European Journal of International Relations. 2019. 25(2):431–57. https://doi.org/10.1177/1354066118781955

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

BIASED PURSUIT GUIDANCE LAW FOR REAR-SECTOR TARGET INTERCEPTION

MILOŠ PAVIĆ
Military Technical Institute, Belgrade, meelos.pavic@gmail.com

SAŠA ANTONOVIĆ
Military Technical Institute, Belgrade

IVAN MARIĆ
Military Technical Institute, Belgrade

Abstract: *This paper presents a Biased Pursuit Guidance Law designed to enforce a terminal target approach from the rear-sector (tail-chase profile). By injecting a time-varying angle bias into the classical pursuit kinematic framework, the proposed guidance law systematically shapes the interceptor's trajectory during initial guidance phase.*

Keywords: *Numerical simulation, biased pursuit, guidance law, rear-sector, tail-chase, proportional navigation.*

1. INTRODUCTION

Classical Pure Pursuit (CPP) commands the interceptor's velocity vector to point directly at the target at all times. While geometrically simple, CPP lacks the trajectory-shaping capabilities needed for tactical operational constraints, such as avoiding forward-aspect defensive countermeasures or aligning an interceptor onboard sensor/payload with a specific quadrant of the target.

Proposed Biased Pursuit Guidance Law (BPGL) achieves rear-aspect target approach by injecting an intentional angular offset from the instantaneous line-of-sight (LOS), and it is used in midcourse phase.

Such an analysis was not found in [1] and [2].

2. ENGAGEMENT GEOMETRY

Consider a planar engagement geometry in an inertial coordinate frame, Fig.1, where:

• M denotes the Interceptor (Missile/UAV) with constant velocity magnitude V_M.
• T denotes the Target.

• R_{TM} denotes the relative range vector between M and T.
• Line-of-Sight (LOS) angle relative to the inertial baseline.

• γ_M denotes the flight-path angle of the interceptor.
• $\varepsilon = LOS - \gamma_M$ denotes the interceptor look-angle (lead angle).

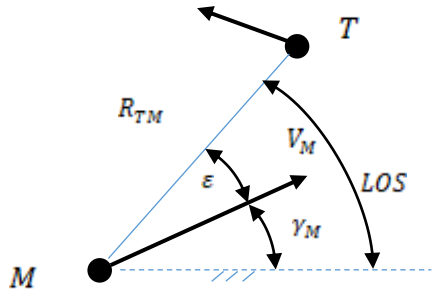


Figure 1. Engagement geometry

3. BIASED PURSUIT GUIDANCE LAW

The Biased Pursuit Guidance Law was defined in form of interceptor lateral acceleration demand by Eq.1, and it was used in first (initial) guidance phase.

$$a_N = Gain * (\varepsilon + Offset); \quad (1)$$

In the second (midcourse) guidance phase, starting when x coordinate of the interceptor approaches the x coordinate of the target to be less than threshold $x_{TM\ Threshold}$, the Offset term is omitted from Eq.1, and in that phase CPP was used to align the velocity vector of the interceptor toward the target position.

In the last (terminal) guidance phase starting when relative distance is less than $R_{TM\ Threshold}$, in order to minimise the miss distance even in case of the maneuvering target, the Proportional Navigation (PN) guidance law with navigation constant 3 was used, Eq.2.

$$a_N = 3 * Vc * \frac{d}{dt}(LOS) \quad (2)$$

Corresponding author: M. Pavić;
E-mail: meelos.pavic@gmail.com
DOI: 10.5937/OTEH26034P

where the relative closing velocity is negative time derivative of the relative range, Eq.3.

$$V_C = -\frac{d}{dt}(R_{TM}) \tag{3}$$

NUMERICAL SIMULATION

Numerical simulation was done in horizontal plane by Fortran code [3] rewritten/translated to Matlab.

Initial value used in numerical simulation was: $Gain = 57.3$ (in order to obtain acceleration command of 1 m/s^2 for lead angle of 1°); $Offset = 50°$; $x_{TM\ Threshold} = 1000\ m$; $R_{TM\ Threshold} = 500\ m$.

The pseudo-code for guidance law is shown in (4)

$$\begin{aligned} &if\ (Rm1 < Rt1 - x_{TM_{Threshold}})\ then \\ &\qquad Offset = 50 * Deg2Rad; \\ &else\ Offset = 0.0; \\ &a_N = Gain * (\varepsilon + Offset) \\ &if\ (Rtm < R_{TM_{Threshold}})\ then \\ &\qquad a_N = 3 * Vc * dLOS\ \%terminal\ guidance \end{aligned} \tag{4}$$

Interceptor initial position was (x=0; y=0), $V_M = 100\ m/s$ toward the target initial position.
Target initial position was (range 10 km, deflection 2 km), $V_T = 50\ m/s$ ahead x-axe to the left.
Interceptor lateral acceleration was simulated as linear first-order system with time constant 0,5 s and limited to maximum value of 3 g.

4. RESULTS

Interceptor and target trajectories in horizontal plane was shown in Fig.1 for two values of $Offset$ (30° and 50°) and two values of $x_{TM_{Threshold}}$ (500 m and 1 km).

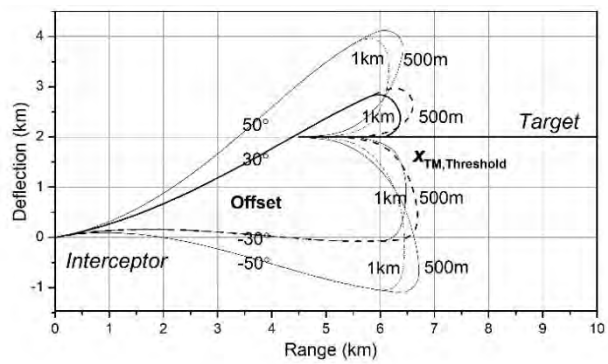


Figure 1. Interceptor and target trajectories in horizontal plane

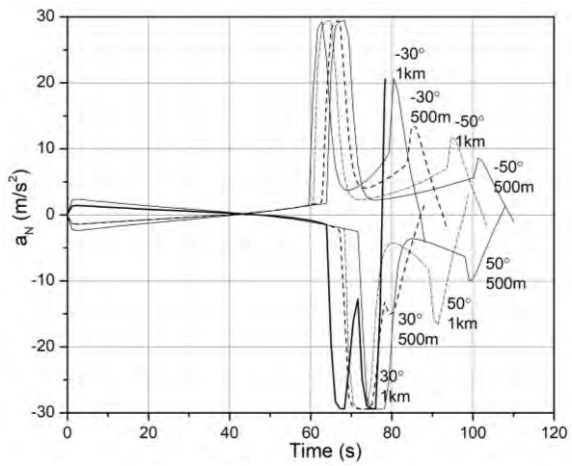


Figure 2. Corresponding missile lateral acceleration

In Fig.2 the corresponding missile lateral acceleration was shown. All terminal missile lateral acceleration values tends to zero, except for $Offset = 30°$ and $x_{TM\ Threshold} = 1\ km$ combination. That is the reason why in that case the final miss distance is significant (0,759 m), as shown in Table 1, Column 5.

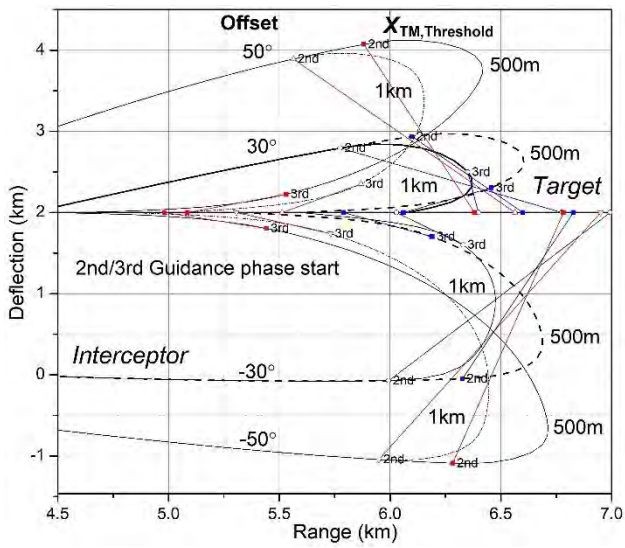


Figure 3. Switching guidance laws trajectory points

In Fig.3 all trajectory points are shown including switching points of different guidance laws (entering the 2-nd and 3-rd guidance phase). Instant corresponding positions of the interceptor and the target are marked and connected together.

One can see that only in case where $Offset = 30°$ and $x_{TM\ Threshold} = 1\ km$, when entering the final (3-rd) terminal guidance phase, the interceptor is still in front of the target (not in rear sector), resulting the larger miss distance.

Total time, energy and final miss distance was shown in Table 1 for all simulated parameter combinations.

Table 1. Total time, energy and final miss distance

Offset [°] (1)	$X_{TM\ threshold}$ [m] (2)	Total time [s] (3)	Total energy [g^2*s] (4)	Final miss [m] (5)
30	500	88,21	8180	0,003
30	1000	78,33	8869	0,759
-30	500	93,60	5126	0,001
-30	1000	88,08	5138	0,003
50	500	107,96	7036	0,001
50	1000	98,80	7097	0,003
-50	500	110,12	5421	0,0002
-50	1000	103,43	5346	0,002

One can see from Table 1 that inducing the larger Offset (Col.1) value results in larger Total time (Col.3). Opposite conclusion stands for $x_{TM\ Threshold}$ (Col.2) where the smaller total time was obtained when this parameter was enlarged.

In case of negative Offsets (Col.1) the Total energy (Col.4) used for guidance was smaller.

5. CONCLUSION

The Biased Pursuit Guidance Law offers an effective, elegant, low-complexity solution for enforcing rear-sector target interception.

By incorporating a bias to the target aspect angle, this guidance law enforces a curvilinear approach path that naturally loops behind the target and guarantees a tail-chase approach across a wide envelope of initial conditions.

This guidance law ensures optimal alignment for tracking systems and payloads. Because it requires minimal computational overhead it can be easily implemented.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] W.R.Chadwick; C.M.Rose.: *A Guidance Law for General Surface Targets,* J.Guidance, 1983, Vol.6. No.6.

[2] D.Cuk; S.Mandic: *Izbor parametara modifikovanog zakona potere samonavodjene letelice*, Naucnotehnicki pregled, 1998, No.3.

[3] P.Zarchan: *Tactical and Strategic Missile Guidance*, 3rd Ed, Progress in astronautics and aeronautics, 1997.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CALIBRATION OF DIAL INDICATORS

DRAGAN LAZIĆ
Technical Test Center, Belgrade, astazu.lazic@gmail.com

ŽAKLINA ČUTOVIĆ
Sloboda ad, Čačak, zaklinacutovic1966@gmail.com

MILOŠ JOVANOVIĆ
Technical Test Center, Belgrade, milosjovanovic1967@gmail.com

Abstract: *In precision manufacturing and quality control, small measurement errors can lead to significant consequences. A deviation of just a few microns may result in rejected parts, rework, production delays, or even non-compliance during audits. This is why dial indicator calibration plays a critical role in maintaining reliable dimensional measurements across industries. The dial indicator (also known as a DTI, clock indicator, dial gauge, probe indicator or comparator) is a precision measuring instrument used for several applications across many industries where precision is necessary, such as manufacturing, packaging, engineering and design. However, without proper calibration, even a well-maintained dial indicator can silently drift out of tolerance. This paper provides a comprehensive overview of dial indicator calibration, covering technical principles, best practices, common hidden errors, and how military calibration laboratory help organizations maintain accuracy and compliance. In this paper describes the influence of the measurement uncertainty to the tolerance, as defined in international standards, and quotes documents on how to estimate measurement uncertainties.*

Keywords: *dial indicator or comparator, calibration, measurement uncertainty*

1. INTRODUCTION

Typical dial gauges (dial indicators) are standard types with 0,01 mm resolution scale, measuring ranges (0-3) mm, (0-5) mm, (0-10) mm and 0,001 mm resolution scale, measuring ranges between (0-1) mm, (0-2) mm, (0-3) mm, (0-5) mm, but there are various models depending on size, purpose, and scale. The dial indicators are used for precise linear measurements (length and diameter) of the runout of rotating elements. Dial gauges with resolution 0,01 mm use for standard dimensional checks it ideal for mass production while resolution 0,001 mm use for high-precision machining and ideal for tight tolerances. Despite significant advancements and availability of sophisticated measurement techniques, such as coordinate measuring machines and optical techniques, dial indicators are still widely used, particularly in the machine industry and research laboratories. This is due to their small size, autonomous operation, satisfactory precision, and availability of accessories that facilitate measurement system configurations. Dial indicators are commonly used for the configuration of precision machining systems, inspection of machine tool accuracy, industrial robot calibration and quality control of batch or one-off production. Measuring accuracy is according to the international standards of Germany DIN 878, DIN 2270, EN ISO 463, and Japan JIS B7503&B7533.

2. WORKING PRINCIPLE OF DIAL INDICATORS

Typically, a dial indicator has a spindle and a clock-face. Any force or pressure applied to the contact point of the spindle will push the spindle up into the body of the indicator or rotate the spindle around an axis.

The dial gauge has some basic components which are a small clock and a plunger (stylus probe). This gauge works when there is very slight upward pressure applied on the plunger that causes it moves upward. The upward movement of the plunger is converted into the rotary movement of an indicator on the dial gauge. Gears in the indicator body magnify the displacement as it moves. The clock face responds to this, showing a reading through a scale and needle on analogue models, and in digital format on digital models. Digital dial indicators replace the mechanical gear train with a linear encoder (optical or inductive) and display displacement on a digital readout.

There are two main dial indicator types: Plunger Dial Indicator and Lever Dial Test Indicator figure 1. Both dial indicator types have similar features while functioning differently. Plunger Dial Indicators (Dial Indicators) use a plunger that moves in a straight line to magnify and convert the force applied to the contact point (at the bottom of the plunger) into a measurement. Plunger dial indicators are also often used in machine setups to monitor changes in the alignment of machine parts and components. The plunger dial indicator needs to be positioned at an angle of 90° to function accurately.

Corresponding author: D. Lazic
E-mail: astazu.lazic@gmail.com

DOI: 10.5937/OTEH26035L

Figure 1. The design of the dial indicator

Lever Dial Test indicators are also known as a lever-type indicator, a DTI or simply a dial test indicator. The bottom of the body houses a pivot with a probe (the lever), its distinguishing feature when visually comparing the two dial indicator types. A test indicator uses a swinging lever arm rather than a linear plunger. The contact stylus pivots on a lever, amplifying small angular displacements into pointer movement. Test indicators access confined spaces and measure at oblique angles inaccessible to standard plunger-type indicators. They are widely used for concentricity measurement on CNC machine tool spindles and fixture alignment.

Besides, the dial gauge is widely used in all types working industries as it has excellent accuracy in measurement. This is because the dial is divided into 100 division and one complete revolution of the indicator corresponds to 1 mmm linear movement of plunger. That means that each division on the dial indicates a movement of 0,01 mm. The basic operating principle of a dial indicator involves a series of precision mechanical components figure 2.

- Contact point (stylus): A hardened, replaceable tip contacts the workpiece surface. Movement of the stylus transmits linear displacement to the instrument.
- Rack and pinion mechanism: The linear motion of the stylus rack is converted to rotational motion by a small pinion gear.
- Gear train amplification: A gear train amplifies the rotation, converting a small linear displacement (0,01 mm) into a large pointer rotation (one full revolution of the large pointer).
- Dial and pointer: A graduated circular dial with a calibrated scale and a rotating pointer displays the amplified displacement measurement. A revolution counter records the number of full rotations for large displacements.
- Return spring: A precision spring returns the stylus to the zero position when contact is removed, ensuring consistent preload on the measured surface.

The measurement resolution (least count) of a standard dial indicator is typically 0,01 mm, with high-precision models achieving 0,001 mm or 0,0001 inch resolution.

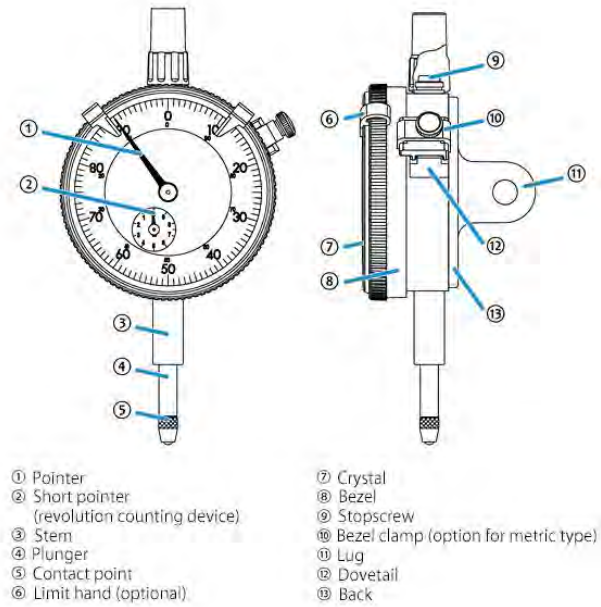


Figure 2. The design of the dial indicator

3. CALIBRATION PROCESS

The calibration is an activity to do periodically by comparing standard quantities between measuring instruments and tester that is in accordance with the measuring instrument. This calibration method was used the procedure to calibrate the dial indicator with a measuring range of (0-25) mm with a resolution of 0,01. The tools and materials used to calibrate the Dial Indicator, namely:

1. Calibration Tester Sylvac, with 0,001 mm accuracy and certificates
2. Weighing scale
3. Thermohygrometer
4. Tissue
5. Petroleum ether/Alcohol
6. A set of instrument screw drivers
7. Work sheet

The measuring surfaces of the calibration device and the gauge measuring surfaces are cleaned using petroleum ether or alcohol. Any grit or other particles are also cleaned from the scale. The surfaces are wiped afterwards using the tissue (or special synthetic cloth). The moving shaft is also cleaned if necessary. The temperature of the tools and dial indicators is stabilised at 20 °C ± 1 °C and humidity of 55 ± 15 % for 5 hours. Keep the instrument and measurement standards under calibration conditions for at least 5 hours prior to the measurements.

After all the tools and materials were available, the next step of the calibration preparation process, which was:

- The dial gauge measure range (0-10) mm resolution 0,01mm compliant with DIN 878 standard is visually checked for any obvious defects. Defects, which are checked are corrosion; missing or worn parts of the scale (on classical scales); missing parts such as screws; and significant scratches or other defects which would impede use.

- Check in advance, whether the dial was functioning properly or not by pressing the end of the sensor, whether it was functioning or not.
- Dial scale is checked for the straightness of the pointer and its distance from the scale. The pointer must be of the same width as the lines on the scale. Lines of the scale must be oriented towards the centre of the dial scale. Resolution must be written on the scale (e. g. 0,01 mm)
- Digital scale is checked for numerical legibility and for any scratches on the display. If the numbers are of poor resolution (checked at display value 88,88) the battery is changed and resolution is checked again. The numbers must be clearly visible in each measuring position.
- The moving part of the calibration device (probe) should be in the lower position. The gauge is fixed in such a way that the pointer is below zero position.
- the measuring range is divided into ten (10) equal increments - ten measuring positions are defined in addition to zero point.
- the probe of the calibration device (which is in contact with the probe of the gauge) is lifted to the position where the pointer of the gauge is exactly in the zero position (the line indicating zero position is covered by the pointer, or a digital display shows 0,00(0) mm).
- the probe of the calibration device is lifted to the first measuring position (positive direction of approaching measuring position); the exact value (e.g. 1,000 mm) should be shown on the display of the device. The deviation from the exact value is read from the dial gauge and is recorded.
- the same procedure is used for other 9 measuring positions.
- after the last deviation is recorded, the probe of the device is lifted for approximately ten scale divisions above the maximum value on the scale and then moved down to the last measuring position (negative direction of approaching measuring position). The deviation of the value indicated by the gauge is recorded again.
- the probe is moved down to other 10 measuring positions (negative direction) and the deviations are recorded.
- the measurement is finished when the last deviation (in zero position) is recorded.

Figure 3. Dial indicator mounted on the calibration tester holder

The following, in DIN 2270 (test indicators) DIN 878 (dial indicators) DIN 879 (dial comparators) defined parameters, will be tested: Span of error f_e, Total span of error f_{ges}, Local span of error f_t, Repeatability f_w, Hysteresis f_u figure 4.

- Span of error f_e is the distance between the ordinates of the highest and lowest points in the deviation diagram either when the plunger is being pressed in (dial indicators, dial comparators) or in the case of excursion of stylus against measuring force (test indicators).
- The total span of error fges is the distance between the ordinates of the highest and lowest points in the deviation diagram when the plunger is being pressed in and let out (dial indicators, dial comparators) or in the case of excursion of stylus against and with the same direction of measuring force (test indicators). The total span of error includes the hysteresis error of the measured value fu.
- Span of error ft is the distance between the ordinates between the highest and lowest points in the deviation diagram, measured for a local measuring span of 10 scale divisions within 0,1 mm when the plunger is being pressed in (dial indicators, dial comparators) or in the case of excursion of stylus against measuring force (test indicators).
- Repeatability fw is a characteristic for measured value fluctuations in measurements of the same measurand within the measuring span and with the same direction of movement of the plunger (dial indicators, dial comparators) or of the stylus (test indicators).

The calibration normally determines f_e, f_{ges}, f_u, f_w, and sometimes f_t.

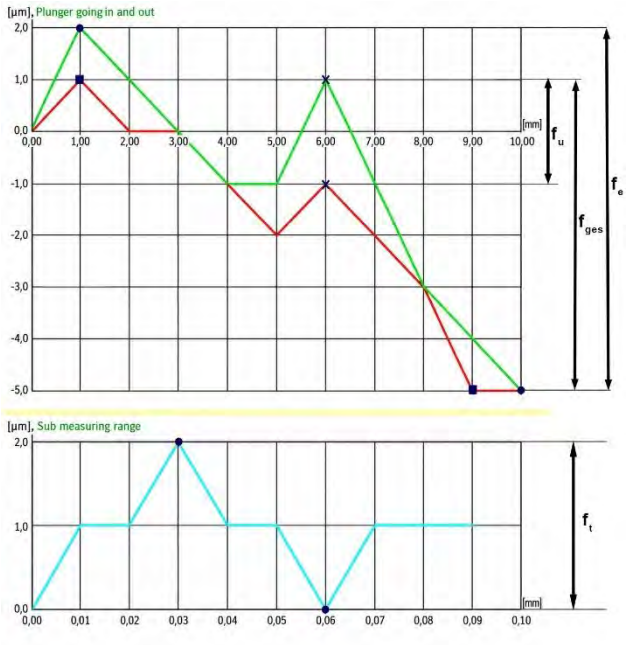


Figure 4. Diagram for the spans of error dial indicator

4. MATHEMATICAL MODEL OF THE MEASUREMENT

The deviation e (result of calibration) is given by the expression:

$$e = l_i \times (1 + \alpha_m \times \theta_m) - l_e \times (1 + \alpha_e \times \theta_e) \quad (1)$$

where:
e - deviation (result of calibration) at 20°C
l_i - indicated value on the dial gauge (reading)
α_m - thermal expansion coefficient of the dial gauge mechanism
θ_m - deviation of the temperature of the dial gauge
l_e - indicated value on the calibration device (plus length of a gauge block where applicable)
α_e - thermal expansion coefficient of the measurement system of the calibration device
θ_e - deviation of the temperature of the measurement system of the calibration device
If new quantities are defined as:

$$\delta\theta = \theta_m - \theta_e \quad (3)$$

$$\delta\alpha = \alpha_m - \alpha_e \quad (4)$$

$\delta\theta$ - temperature difference between the dial gauge and the calibration device
$\delta\alpha$ - difference in their thermal expansion coefficient
then equation (1) gets the following form:

$$e = l_i \times (1 + \alpha_m \times \theta_e + \alpha_m \times \delta\theta) - l_e \times (1 + \alpha_m \times \theta_e - \delta\alpha \times \theta_e) \quad (5)$$

The sensitivity coefficients are the partial derivatives of the model:

$$e = f(l_i, \alpha_m, \theta_e, \delta\theta, l_e\delta\alpha) \quad (6)$$

$$c_{l_i} = \frac{\partial e}{\partial l_i} = 1 + \alpha_m \times \theta_e + \alpha_m \times \delta\theta =\approx 1$$

$$c_{\alpha_m} = \frac{\partial e}{\partial \alpha_m} = l_i \times (\theta_e + \delta\theta) - l_e \times \theta_e$$

$$c_{\delta\theta} = \frac{\partial e}{\partial \delta\theta} = l_i \times \alpha_m$$

$$c_{l_e} = \frac{\partial e}{\partial l_e} = -1 + \alpha_m \times \theta_e - \delta\alpha \times \theta_e =\approx 1,$$

$$c_{\delta\alpha} = \frac{\partial e}{\partial \delta\alpha} = l_e \times \theta_e$$

where is $l_i = 5\ mm,\ l_e = 5{,}003\ mm,\ \alpha_m = 11{,}5 \times 10^{-6}\ °C^{-1},\ \delta\alpha = 1 \times 10^{-6}\ °C^{-1}, \delta\theta = 0{,}2\ °C,\ \theta_e = 1\ °C$
Estimation of standard uncertainties of the input quantities:

- $u(l_i)$, resolution of dial gauge, d=0,01 mm, Type B from resolution, half pass of d, rectangular distribution $u(l_i)=d/\sqrt{3}$, $u(l_i) = \frac{d}{\sqrt{12}} = 2{,}89$ µm.
- $u(\alpha_m)$, manufacturer's specification or literature, Type B, rectangular distribution, $\alpha_m=11{,}5\times10^{-6}$ °C^{-1} ±1×10^{-6} °C^{-1} (typical steel value) $u(\alpha_m) = 1 \times \frac{10^{-6}}{\sqrt{3}} = 0{,}577 \times 10^{-6}$ °C^{-1}.
- $u(\theta e)$, calibration of temperature sensor and temperature stability, combination of sensor calibration, resolution and stability. Temperature deviation at 20°C, θe=±1°C, rectangular distribution $u(\theta_e)=\theta_e/\sqrt{3}=0{,}577$°C
- $u(\delta\alpha)$, difference between thermal expansion coefficients $\delta\alpha=\pm1{,}5\times10^{-6}$ °C^{-1}, rectangular distribution $u(\delta\alpha)=\delta\alpha/\sqrt{3}=0{,}866\times10^{-6}$ °C^{-1}
- $u(\delta\theta)$, temperature difference between dial gauge and reference. Estimated from thermal equilibrium and sensor uncertainty δθ = ±0,2°C, rectangular distribution, $u(\delta\theta)= \delta\theta/\sqrt{3}=0{,}115$°C
- $u(l_e)$, calibration device, expanded uncertainty from certificate $u(l_e) = 0{,}06 + 5 \times 10^{-4} \times l$, l (mm).

- **Table 1**. Uncertainty Budget

Uncert source	Estimate	Distrib.	Stand. uncert. u(xi)	Sensitivity coeff.	Uncert. Contrib.
Uncertainty of the reading of the result u(li)	0,01 mm	Rectan.	2,89 µm	1	2,89
Uncertainty of the thermal expansion coefficient u(αm)	11,5×10−6 °C−1	Rectan.	0,577×10−6 °C^{-1}	0,001×10^6 µm x°C	6,640x10^{-3}
Uncertainty of the thermal expansion coefficient difference u(δα)	1,5·10-6°C^{-1}	Rectan.	0,866×10^{-6} ° C^{-1}	0,005×10^6 µm x°C	4,330x10^{-3}
Uncertainty of the temperature deviation u(θ_e)	1°C	Rectan.	0,577 ° C^{-1}	0,0075×10^6 µm x°C^{-1}	4,330x10^{-3}
Uncertainty of temperature difference u(δθ)	0,2°C	Rectan.	0,115 °C	0,0575 µm x°C^{-1}	6,640x10^{-3}
Uncertainty of indication of the device u(l_e)	0,06+5x10^{-4}xl	Normal	0,032 µm	1	6,250x10^{-2}
Combined standard uncertainty (uc)					2,887
Expanded uncertainty *U* (k=2) µm					5,78

Calculation (estimation) of standard uncertainties of influence values is given in Table 1 for the equipment, methods and conditions used in our laboratory.

For measurement point 5 mm measurement uncertainty is 5,78 µm.

Complete result l_i'= l_i ±U = 5 mm ± 0,00578 mm.

5. RESULT AND DISCUSSION

Measured displacement between the pointer positions 0 mm (initial position) and 5 mm (end position):
When the pointer position is l_i = 5,00 mm, the standard

device indicates the measured displacement l_e = 5,003 mm.

The deviation of the dial gauge indication Δl_i from the conventional value l_e of the standard device $\Delta l_i = l_e - l_i$ is 3 µm. In practice, corrections are not common for this type of dial gauge so that the systematic error must be considered as an uncertainty contribution in the uncertainty budget. In metrology and dial gauge calibration (under standards like DIN and ISO 463), f_e=15 µm (span of error/deviation span) is the maximum error when moving the plunger in a single direction, while f_{ges}=17 µm (total deviation span) is the overall error range combining both directions, which includes mechanical hysteresis. Accordingly, the conventional value of the measurement result can be expected in the range between 4,99 mm and 5,006 mm with a confidence level of 95.45%. This applies to the 5 mm measuring point only. If the deviation error and measurement uncertainty are analyzed, the measurement is in the manufacturer's specification.

6. CONCLUSION

In a process of calibrating a measuring instrument, it is influenced by several factors such as the uncertainty of the measuring instrument, the uncertainty of the calibration tool, the standard deviation, changing in temperature, general uncertainty and the combined uncertainty. Therefore, from all the values of these factors, the uncertainty value can be obtained. Based on the calibrations that have been made, several conclusions can also be drawn, which is:

- Calibration is a standard procedure in factories/companies, especially those engaged in production, because it can control the quality of the equipment used, and can show whether the tool is feasible or not to be used.
- The obstacle experienced during dial Indicator calibration is the concentration of the operator performing the calibration, which can affect the results of readings on the dial and on standard equipment. There are also some cases where the measuring capacity and reference standards are not suitable when trying to calibrate.

References

[1] Pravilnik o metrološkim uslovima za mehaničke merne satove s vrednossću podeljka 0,01 mm i sa mernim hodom do 10 mm Sl SFRJ br 5 1985

[2] Metrološko uputstvo za pregled mernih satova MUP.D-9/1, Službeni glasnik Saveznog zavoda za mere i dragocene metale br 3/87.Dial indicators for linear measurements, ASME B89.1.10M-2001, Revision of ASME/ANSI B89.1.10M-1987 (R1995).

[3] Pravilnik o metrološkim uslovima za merne satove komparatere MUS .01MSO0203-01 SL SRJ br. 26 od 1995

[4] ITTC – Recommended Procedures and guidelines, Calibration of Dial Gauges, 2017

[5] Quality Management in the Bosch Group Technical statistics Booklet 8 – Measurement Uncertainty Edition 06/2015

[6] https://www.farnell.com/datasheets/55344.pdf

[7] https://www.mitutoyo.com/webfoo/wp-content/uploads/E11003_7_Quick_Guide_DB.pdf

[8] Calibration of Dial Indicator Using Calibration Tester with JIS B 7503 Standard, Journal of Ocean, Mechanical and Aerospace, July 30, 2021

[9] Procedure for calibrating dial indicators, UNITED STATES DEPARTMENT OF THE INTERIOR BUREAU OF RECLAMATION, USBR 1007-89

[10] SOP 8 - CALIBRATION OF DIAL GAUGES, LABORATORY FOR PRODUCTION MEASUREMENT, Faculty of Mechanical Engineering Smetanova 17, 2000 Maribor, Slovenia, 2003.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

TOTAL QUALITY INTEGRATED IN NEW AMMUNITION PRODUCTS IN AD "PRVI PARTIZAN" (ACCORDING TO ISO/TR 14062 - 2002)

SVETOLIK SPASENIĆ
AD „Prvi partizan", Užice, svetolik.spasenic@prvipartizan.com

***Abstract**: This papar presents a global market trend indicating that input requirements for product design and development must, in addition to quality and safety demands, integrate requirements at an early design stage ensuring that the product is harmless to nature and the environment (achieving total quality).*

Aware of such user needs and social demands regarding the human right to a healthy environment, "Prvi partizan " AD has designed and implemented the production of a new generation of products – ammunition with TMJ (Total Metal Jacket) type bullets.

Following the firing of a round and upon leaving the muzzle, such bullets do not release harmful lead emissions into the environment.

Keywords: *new ammunition products, environmental requirements, total product quality*

1. INTRODUCTION

The eighty years of experience of "Prvi partizan" AD in the design and production of high-quality and safe ammunition products has recently been enriched by the design of products that also integrate eco-friendly and environmental requirements – total quality products. In SRPS ISO 14004 (1997), Appendix A2 of the Business Charter for Sustainable Development of the International Chamber of Commerce (ICC), Principle 6 is defined as follows: "To develop and provide products or services which have no undue environmental impact and are safe in their intended use, which are efficient in their consumption of energy and natural resources, and which can be recycled, reused, or disposed of safely."

Aware that every product exerts some impact on the environment throughout its life cycle (starting from raw material procurement, product realization, storage, use, recycling, and waste disposal), we have identified a significant environmental impact – harmful lead emissions into the atmosphere upon firing pistol and revolver ammunition assembled with FMJ (full metal jacket) type bullets. We have decided to include environmental requirements for our products at the earliest stage of product design, i.e., in the input requirements for the design and development of new products, in order to achieve the overall – total quality of our products. The result of the design and development process was TMJ (total metal jacket) or CMJ (complete metal jacket) type bullets, as they are called by a well-known global manufacturer in this field (Israel Military Industries Ltd. IMI). In this way, we have fulfilled Principle 6 of the aforementioned Charter.

2. GLOBAL MARKET REQUIREMENTS

"Prvi partizan" AD has been present on the global market with its products for years, thanks to the fact that it has adopted the technical standards of these markets (ANSI/SAAMI for North American, and C.I.P. for European and other markets) in the field of commercial weapons and ammunition as its own and applies them unconditionally. In this way, technical barriers to mutual trade exchange have been eliminated, but new requirements are emerging, particularly from developed countries, stating that the environment must be preserved from pollution and harmful impacts during the product's life cycle. Our policy and strategy have always been that all user requirements, whether explicitly stated or unstated but implied, must be fulfilled and their expectations exceeded. At the same time, all legal requirements of the market to which our product is distributed must be properly understood and met. Thus, the top management of the company has adopted a policy that all user requirements, including environmental ones, must be fulfilled if we want to ensure survival and market share on the global market, especially the market of developed countries, where legislation on human health and environmental protection has been intensively enacted over the past 20 years.

3. IDENTIFICATION OF ENVIRONMENTAL IMPACT

The process of firing a cartridge begins when the firing pin strikes the primer. This strike ignites the primer mixture contained within the primer. The flame from the primer is transferred to the powder charge, after which the combustion of the gunpowder takes place. This generates a very high temperature inside the barrel of the weapon

Corresponding author: S. Spasenić;
E-mail: svetolik.spasenic@prvipartizan.com

DOI: 10.5937/OTEH26036S

(higher than 2000°C), meaning that the core of an FMJ type bullet (the core is made of a lead-antimony alloy where the antimony content is less than 3%) is exposed to a temperature greater than the vaporization point of lead. The melting and vaporization points of lead and antimony are given in Table 1.

Table 1. The melting and vaporization points Pb and Sb

Component of the bullet core	Melting point (°K)	Vaporization point (°K)
Lead (Pb)	600,3 (327,3° C)	2013 (1740°C)
Antimony (Sb)	903,5 (630,5°C)	1713 (1440°C)

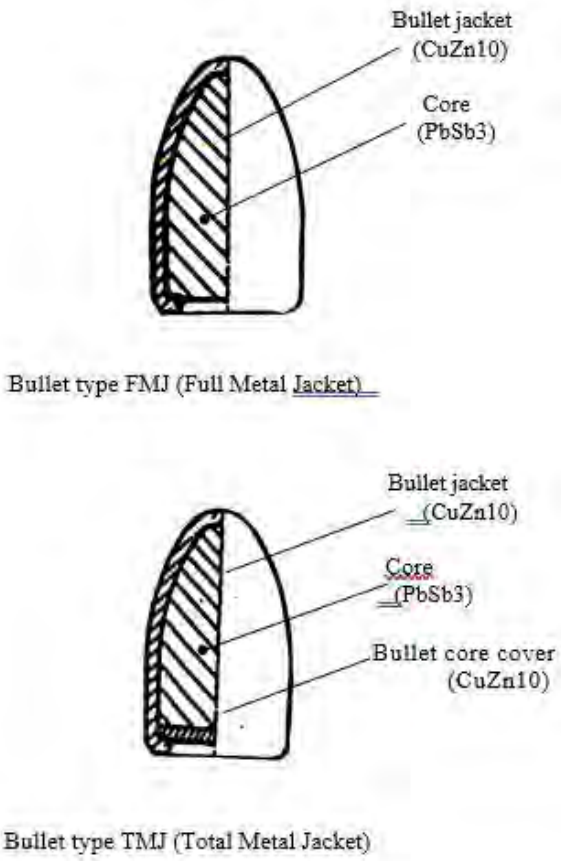


Figure 1.

The consequence of exposing the bullet core to such an extremely high temperature (even in a very short period of time – a few hundredths of a second) is a harmful emission of lead into the atmosphere upon firing the cartridge, i.e., after the bullet leaves the muzzle of the weapon, as well as a harmful impact on the soil at the point of impact due to direct contact between the soil and the lead.

The effect of harmful lead emissions into the atmosphere is particularly intensified when ammunition is fired in poorly ventilated or enclosed shooting ranges, as lead, being a heavy metal, is difficult to remove via ventilation or airing. Furthermore, with every firing, the lead deposited on the walls and floors of the test ranges in the form of microparticles is kicked up and swirled into the atmosphere, making work (or training) very difficult and hazardous to health. It should also be noted here that lead in the form of a fine powder, i.e., microparticles, becomes flammable in the presence of oxygen – so-called **pyrophoric lead** – presenting an additional potential hazard: the risk of fire.

We can notice how "heavy" a metal lead is, from its position in Mendeleev's periodic table of elements, where its atomic number is 82, with 82 protons, 82 electrons, and 125 neutrons in its atom; its relative atomic mass is $A = 207.2$ (for example, uranium has the atomic number 92 with a relative atomic mass of $A = 238$).

After identifying this type of environmental impact when using cartridges with FMJ-type bullets, it became clear that another group of limiting factors – environmental factors – must be included in the product design input requirements. These factors, regardless of everything else, must be fulfilled within the **process of designing total product quality**.

4. DESIGN PHASES OF A CARTRIDGE WITH A TMJ TYPE BULLET (TOTAL METAL JACKET)

By incorporating environmental factors into the input requirements for the design of ammunition products, TMJ-type bullets – total quality bullets – were created. The entire process of designing a cartridge with a TMJ-type bullet went through the following phases:

4.1. Planning phase

In this phase, it was planned which specific calibers would be included in the design process and with which input requirements. The product life cycle was also considered, along with identifying the specific stages of the life cycle where the product has a significantly harmful impact on the environment.

4.2. Conceptual design phase

In this phase, the final product design was defined, the manufacturing technology was engineered, and the technical documentation (TD) for production, as well as the technical documentation for testing and inspection, was established.

4.3. Prototype development phase

In this phase, the final product design was defined, the manufacturing technology was engineered, and the technical documentation (TD) for production, as well as the technical documentation for testing and inspection, was established.

4.4. Prototype testing phase (verification and validation)

In this phase of prototype testing (verification and validation tests), in addition to all other quality

requirements (from laboratory to experimental and operational), the environmental performance of the product was also inspected. In all tests, the bullet remained compact and without cover separation from the bullet core, which guaranteed the elimination of harmful lead emissions from the lead core at the initial part of the bullet trajectory; as well as, after the bullet fall, direct contact of the soil with lead from the bullet core. The development was successfully completed.

4.5. Phase of serial production and market launch of the product

The product has entered serial production and has been delivered to the user, securing its rightful place as a new product in our ammunition catalog for year 2012.

4.6. Phase of product evaluation by users

Users - buyers have confirmed that we have met all their expectations, to the mutual satisfaction for both parties, we have achieved high overall quality of the product with reliability and safety in use, and in addition, the product is harmless to nature and the environment.

5. PRODUCTS WITH BULLET TYPE TMJ

In 2002, AD "Prvi partizan" developed the following new ammunition products by integrating environmental protection requirements into design input requirements (ecological requirements), somehow just at the time when the international technical recommendation from the group of standards on environmental management ISO/TR 14062 was introduced - Integrating of environmental aspects into product design and product development.

Table 2. Cartridge with TMJ bullet type produced by „Prvi partizan" AD

No.	Caliber	Bullet type	Bullet mass (grain)	Bullet mass (g)
1.	9 mm Luger	TMJ	115	7,45
2.	9 mm Luger	TMJ	124	8,0
3.	38 Special	TMJ	158	10,25
4.	40 Smith &Wesson	TMJ	180	11,7
5.	10 mm AUTO	TMJ	180	11,7
6.	9 mm x 21	TMJ	115	7,45
7.	9 mm x 21	TMJ	124	8,0
8.	45 AUTO	TMJ	230	14,85

In the process of producing bullet type TMJ (CMJ), it is spent up to 10-15% more time per unit, which to some extent increases the cost of production, but therefore **achieves a significant advantage over the competition in the market**. This gives the company potential benefits such as:

- gaining a vision of developing and delivering products that do not have harmful impact on the environment and are safe for their intended use
- this ensures that users have an obligation to protect the environment while satisfying their wishes and needs
- maintaining good relations with the public and the wider social community
- increasing reputation and market share
- meeting the criteria for certification according to SRPS ISO 14001
- facilitating the acquisition of permits and possibilities for placement in developed countries
- improving relations with the competent ministry for environmental protection.

By reviewing the catalog of the renowned ammunition manufacturer "Israel Military Industries ltd (IMI)", we have noticed that they have produced TMJ bullets in the same calibers as we did, except for the calibers: 40 Smith & Wesson, 10 mm AUTO and 45 AUTO. Our advantage is obvious.

Table 3. List of cartridges with bullet type CMJ manufactured by „Israel Military Industries (IMI)"

No.	Caliber	Bullet type	Bullet mass (grain)	Bullet mass (g)
1.	9mm Luger	CMJ	115	7,45
2.	9 mm Luger	CMJ	124	8
3.	38 Special	CMJ	125	8,1
4.	9 mm x 21	CMJ	115	7,45
5.	9 mm x 21	CMJ	124	8

6. CONCLUSION

Product designers and engineers at the company "Prvi Partizan A.D." are facing new challenges in integrating all previously established requirements related to product and service quality and safety, along with current requirements for environmental and nature protection throughout every stage of the product or service life cycle, with the aim of achieving total quality.

It is essential to study and implement the standards from the SRPS ISO 14000 series, which relate to products and services, as follows:

– SRPS ISO 14020 (2001): Environmental labels and declarations – General principles
– SRPS ISO 14024 (2001): Environmental labels and declarations – Type I eco-labelling – Principles and procedures
– SRPS ISO 14040 (2000): Environmental management – Life cycle assessment – Principles and framework
– SRPS ISO 14041 (2000): Environmental management – Life cycle assessment – Defining the objective and scope and inventory analysis
– SRPS SO 14042: Environmental management – Life cycle assessment – Life cycle impact analysis
– SRPS ISO 14043: Environmental management - Life cycle assessment – Life cycle interpretation
– SRPS ISO Guide 64 (2000) Guidelines for the inclusion of environmental aspects in product standards
– ISO/TR 14062 (2002) – Environmental management – Integrating environmental aspects into product design and development.

It would be beneficial for national legislation to introduce incentives for organizations that design and deliver products with no adverse impact on nature and the environment — products of total quality. Such measures would quickly produce positive results.

These efforts would facilitate the integration of the Serbian economy into international trade and strengthen its position in the demanding global market, influenced by numerous trade limitations, regulatory requirements, and restrictive economic policies.

References

[1] Cvetković M. Internal ballistics I, Zagreb, 1981.
[2] Jakšić B. Integrating environmental aspects into product design and development.
[3] Krstić I. National standardization system in the Republic of Slovenia
[4] ISO/TR 14062 (2002) – Environmental management– Integrating environmental aspects into product design and development
[5] SRPS ISO 14001 (1997) – Environmental management systems – Specification with guidance for implementation
[6] SRPS ISO 14004 (1997) – Environmental management systems – General guidelines for principles, systems and procedures

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

OPTIMIZATION OF CHASSIS GEOMETRY FOR INCREASING THE PRODUCTION CAPACITY OF 4X4 VEHICLE PLATFORM

MIHAJLO VASILJEVIĆ
Zastava TERVO, Kragujevac, mihajlo.vasiljevic@zastavatervo.rs

BOGDAN ŽIVKOVIĆ
Zastava TERVO, Kragujevac, bogdan.zivkovic@zastavatervo.rs

JOVANKA LUKIĆ
Faculty of engineering University of Kragujevac, Kragujevac, lukicj@kg.ac.rs

Abstract: *The optimization of chassis geometry represents a key step in the development of modern modular 4×4 vehicle platforms. The manufacturing of chassis with complex geometries requires the use of specialized tooling, involves potential deformations due to welding, and demands a complex logistical framework, all of which significantly increase production time and costs. These requirements further limit production flexibility and may slow down adaptation to diverse operational demands.*

In the military sector, where rapid response and adaptability to various missions are essential, modularity and shortened production cycles—along with minimal reliance on specialized tools—are of critical importance. For this reason, a chassis concept with simplified geometry and standardized structural elements has been developed, enabling rapid configuration of different superstructures. The entire assembly process is significantly simplified, as only a single type of longitudinal beam needs to be produced, while the wheelbase can be easily adjusted according to specific application requirements, owing to a dedicated connection method of the suspension subassembly. This approach enables more efficient production, easier maintenance, and greater operational flexibility of the vehicle.

Keywords: *vehicle modularity, increased production capacity, rapid assembly*

1. INTRODUCTION

Modern wheeled military vehicles are expected to provide high mobility, adaptability and operational flexibility while supporting a wide range of mission-specific configurations. To meet these demands, vehicle manufacturers are increasingly adopting modular platform concepts that enable multiple vehicle variants to be developed from a common structural base[1,2]. Such an approach can improve production flexibility [1,2], while modularity in military vehicle design may also provide benefits related to adaptability, life-cycle costs, maintenance and logistics [1].

The chassis is the primary load-bearing structure of a vehicle and has a significant influence on its structural performance, manufacturability and operational capabilities [3,4]. Conventional ladder-frame chassis remain the most widely used solution for military and off-road vehicles due to their robustness and versatility[3]. However, their production often involves complex frame geometries, numerous structural components, dedicated manufacturing fixtures and extensive welding operations. These factors increase production time and cost, reduce manufacturing flexibility and may introduce welding-induced distortions that require additional corrective operations[3,4].

Current trends in military vehicle development therefore focus on increasing modularity, component standardization and production simplification [1,3]. Reducing the number of unique structural components while enabling different vehicle configurations from a common chassis can significantly improve manufacturing efficiency and logistical support, particularly in low-volume military production[1,3].

This paper presents the development of a novel modular chassis concept for a multi-purpose 4×4 vehicle platform. The proposed design is based on simplified chassis geometry and standardized structural elements, aiming to reduce manufacturing complexity, minimize the need for dedicated tooling and improve production flexibility while maintaining the adaptability required for various military applications.

2. REVIEW OF CONVENTIONAL CHASSIS ARCHITECTURES

The chassis is the primary structural component of a vehicle, responsible for supporting the powertrain, suspension, body and payload while ensuring adequate strength and stiffness under various loading conditions. Over the years, several chassis architectures have been

Corresponding author: M.Vasiljević
E-mail: mihajlo.vasiljevic@zastavatervo.rs
DOI: 10.5937/OTEH26037V

developed, each offering specific advantages depending on the intended vehicle application. The most commonly used structural concepts include ladder frame, backbone, monocoque and space frame chassis. Their design characteristics differ in terms of structural efficiency, manufacturing complexity, modularity and suitability for off-road and military applications[3].

2.1 Ladder frame chassis

The ladder frame is the most widely adopted chassis architecture for commercial trucks, off-road vehicles and military platforms. It consists of two longitudinal side members connected by a series of cross members, forming a robust load-bearing structure capable of supporting different body configurations. Its relatively simple construction, high load-carrying capacity, ease of maintenance and compatibility with various suspension systems have made it the preferred solution for low- and medium-volume production of military vehicles. Furthermore, the body-on-frame concept enables different mission-specific superstructures to be mounted on a common chassis platform, providing a certain degree of versatility. Despite these advantages, conventional ladder-frame chassis often require multiple frame variants, dedicated manufacturing fixtures and extensive welding operations, which increase production complexity and cost[3].

2.2 Backbone chassis

Backbone chassis utilize a central longitudinal structural member that carries the primary loads and connects the front and rear suspension assemblies. This architecture provides high torsional and bending stiffness, while drivetrain components can be enclosed and protected within the central structure [4]. Although less commonly used than conventional ladder-frame configurations, backbone chassis are still employed in certain specialized off-road and military vehicles. One notable example is the Tatra chassis concept, which combines a central backbone tube with independently suspended swinging half-axles [5].

2.3 Monocoque structure

In monocoque vehicles, the body itself serves as the primary load-bearing structure, eliminating the need for a separate frame. This concept offers significant weight reduction, improved crash performance and high structural efficiency, making it the dominant solution in passenger vehicle design[4]. Nevertheless, monocoque structures are less suitable for heavy-duty military applications, where high payload capacity, ease of repair and adaptability to multiple vehicle configurations are essential requirements[4].

2.4 Space frame structure

Space frame chassis consist of interconnected tubular or closed-section members arranged to achieve high stiffness with relatively low structural mass. Owing to their excellent strength-to-weight ratio, these structures are frequently used in racing vehicles and certain specialized military applications [6]. Recent studies have shown that space frame architectures can provide improved structural performance under high dynamic loading conditions [7] while offering greater design flexibility and modularity than conventional monocoque structures. However, their complex geometry, numerous joints, demanding fabrication processes and extensive welding requirements significantly increase manufacturing time and cost, limiting their suitability for cost-effective production of modular military vehicle platforms [4].

2.5 Limitations of conventional chassis designs

2.5.1 Manufacturing complexity

Conventional chassis architectures, particularly ladder-frame designs used in military vehicles, often require numerous structural components with different geometries and dimensions. Multiple longitudinal rails, cross members, reinforcement brackets and connection elements increase the number of manufacturing operations, assembly stages and quality control procedures. As vehicle families expand to include multiple wheelbase and payload variants, the number of unique structural components further increases, leading to greater production complexity and higher manufacturing costs [8].

2.5.2 Dedicated manufacturing tooling

Conventional chassis production frequently relies on dedicated manufacturing fixtures, positioning tools and welding jigs specifically designed for individual chassis variants [9]. The introduction of additional wheelbase configurations generally requires new tooling, increasing investment costs and extending production preparation time. Such an approach reduces manufacturing flexibility, particularly in low-volume military vehicle production where rapid adaptation to changing operational requirements is essential.

2.5.3 Welding and assembly challenges

The large number of welded joints required in conventional chassis structures increases manufacturing time and may introduce residual stresses and geometric distortions [10]. These effects often require additional inspection and corrective operations to ensure dimensional accuracy and structural integrity. Furthermore, extensive welding complicates repair and maintenance throughout the vehicle life cycle, particularly for vehicles operating under demanding field conditions.

2.5.4 Limited modularity

Although conventional ladder-frame chassis allow different vehicle bodies to be mounted on a common platform, their modularity is often limited by the need for different frame geometries and wheelbase-specific structural components. Consequently, each vehicle variant frequently requires modifications of the primary load-bearing structure, additional manufacturing documentation and dedicated production resources. These

limitations reduce production efficiency and increase logistical complexity throughout the vehicle life cycle[3]. A comparison of the main characteristics, advantages and limitations of conventional chassis architectures is presented in Table 1.

Table 1. Limitations of conventional chassis architectures

Limitation	Manufacturing impact	Operational impact
Multiple structural components	Higher production cost	More spare parts
Dedicated tooling	Longer production preparation	Reduced flexibility
Extensive welding	Distortion and longer assembly	More difficult maintenance
Wheelbase-specific designs	Additional chassis variants	Increased logistics complexity

These limitations highlight the need for a simplified and highly modular chassis concept capable of accommodating multiple vehicle configurations while minimizing structural complexity, production tooling and manufacturing effort.

3. DESIGN REQUIREMENTS

The design of a military vehicle chassis should satisfy both operational and manufacturing requirements. In addition to providing sufficient structural strength and reliability under various operating conditions, the chassis should enable efficient production, simplified maintenance and adaptability to different vehicle configurations. Based on the review of conventional chassis architectures and the requirements of modern military vehicle platforms, a set of operational and manufacturing requirements was established to guide the development of the proposed chassis concept [1,3].

3.1 Operational requirements

The proposed chassis concept was developed to satisfy the following operational requirements:

– adaptability to different mission profiles;

– compatibility with various mission-specific superstructures;

– accommodation of different wheelbase configurations;

– simplified maintenance and servicing throughout the vehicle life cycle.

3.2 Manufacturing requirements

In addition to operational performance, the proposed chassis should facilitate efficient and economical manufacturing. Therefore, the following manufacturing requirements were considered during the design process:

– simplified structural geometry;

– reduced dependence on dedicated manufacturing fixtures and tooling;

– shorter production and assembly time;

– reduced production and life-cycle costs.

4. PROPOSED MODULAR CHASSIS CONCEPT

The proposed chassis concept was developed to satisfy the operational and manufacturing requirements presented in the previous section while preserving the structural characteristics required for multi-purpose military vehicles. The development focused on simplifying the structural configuration, reducing the number of unique components and increasing manufacturing flexibility without compromising the adaptability of the platform.

Unlike conventional chassis designs, where different vehicle variants often require dedicated structural components and manufacturing tooling, the proposed concept is based on a modular architecture composed of standardized structural elements. The chassis was designed to enable straightforward manufacturing and assembly using conventional fabrication processes and widely available equipment, allowing production to be established with minimal dependence on specialized tooling. Such an approach is particularly suitable for low-volume military vehicle production and situations requiring rapid manufacturing or repair under field conditions.

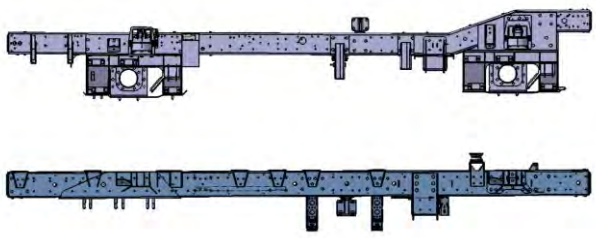


Figure 1 - Comparison of the baseline chassis and the proposed chassis concept

The proposed chassis concept incorporates several design features aimed at improving manufacturability and operational flexibility. These include a geometrically optimized chassis structure, modularly symmetrical longitudinal members, standardized cross members and an adjustable wheelbase concept. Each of these design features is described in the following sections.

4.1 Simplified structural design

The proposed chassis was developed with the objective of simplifying the structural configuration while preserving the strength, functionality and adaptability required for multi-purpose military vehicle platforms. The design process focused on reducing the number of structural components, simplifying manufacturing operations and improving assembly efficiency without compromising the structural performance of the chassis.

In addition to optimizing the geometry of the primary load-bearing structure, an important design decision was the selection of an independent suspension system from a different manufacturer incorporating an integrated suspension subframe. Compared to the previously used

solution, in which the suspension mounting structure consisted of multiple individual components assembled directly onto the chassis, the integrated subframe significantly simplified the overall structural configuration. This approach considerably reduced the number of structural components, assembly operations and manufacturing time required for chassis production.

Since a portion of these improvements results from the adoption of the integrated suspension subframe rather than solely from the redesigned chassis structure, the comparison presented in this study is performed at two levels. First, the primary chassis structures are compared independently of the suspension subframes in order to evaluate the influence of the proposed structural redesign itself. Subsequently, the complete chassis assemblies, including the suspension subframes, are compared to assess the overall manufacturing and assembly benefits of the proposed vehicle platform. This approach provides a balanced and objective evaluation by distinguishing improvements resulting from the chassis redesign from those associated with the selected suspension architecture.

The simplified structural design established the foundation for the remaining design features of the proposed chassis concept, including standardized longitudinal members, modular cross members and an adjustable wheelbase configuration, which are described in the following sections.

4.2 Symmetrical interchangeable longitudinal members

The longitudinal side members are the primary load-bearing elements of the chassis and therefore have a significant influence on manufacturing complexity, logistics and assembly. Conventional chassis designs typically employ separate left- and right-hand longitudinal members, resulting in a larger number of unique structural components and increasing the complexity of manufacturing, inventory management and production planning.

To simplify the production process and improve component standardization, the proposed chassis incorporates interchangeable longitudinal members. A single standardized component was developed that can be installed on either side of the chassis, eliminating the need for dedicated left- and right-hand parts. This approach reduces the number of unique structural components, simplifies inventory management and production planning and facilitates maintenance and replacement throughout the vehicle life cycle.

An additional advantage of the proposed design is the elimination of one of the most common manufacturing errors associated with chassis production. In conventional chassis manufacturing, separate left- and right-hand longitudinal members are produced from different unfolded geometries. Consequently, bending a component according to the drawing of the opposite side results in an unusable part, leading to material waste, additional manufacturing costs and production delays.

In the proposed design, both longitudinal members are manufactured from an identical, geometrically symmetrical unfolded geometry. Owing to this symmetry, the finished component remains suitable for installation regardless of the bending direction. Consequently, a correctly manufactured part cannot become unusable due to an incorrect interpretation of the left- or right-hand configuration, as shown in Figure 2. The possibility of producing an incorrect longitudinal member is therefore eliminated by the structural design itself rather than by manufacturing procedures or quality control.

Besides reducing the number of unique structural components, the proposed solution increases manufacturing reliability, simplifies quality control and inventory management, reduces material waste and minimizes the potential for human error during fabrication.

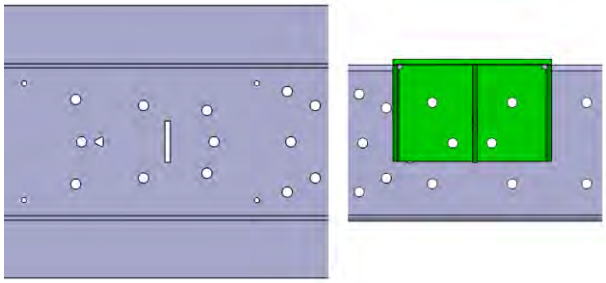


Figure 2 - Symmetrical interchangeable longitudinal member

The proposed symmetrical design results in several unused openings when the longitudinal member is installed on one side of the chassis, as these openings are intended for components that become active when the same member is installed on the opposite side, as shown in Figure 2. Although this represents a minor increase in the number of unused features, it is considered an acceptable design trade-off in view of the significant manufacturing and logistical benefits achieved.

The use of a single interchangeable longitudinal member enables identical spare parts to be used regardless of the installation side, eliminating the need to stock separate left- and right-hand components. Furthermore, in field conditions, replacement can be carried out without any structural modifications or additional machining. An additional advantage is that longitudinal members recovered from damaged or disabled vehicles can be directly reused as spare parts for other vehicles, providing valuable logistical support in emergency situations where replacement components may not be readily available.

4.3 Modular cross members

In order to increase the torsional stiffness of the chassis while preserving its modular character, a cross member manufactured by laser cutting and simple bending operations was developed. This solution enables a relatively simple production process and avoids the need for complex forming tools. Installed cross members are shown in Figure 3.

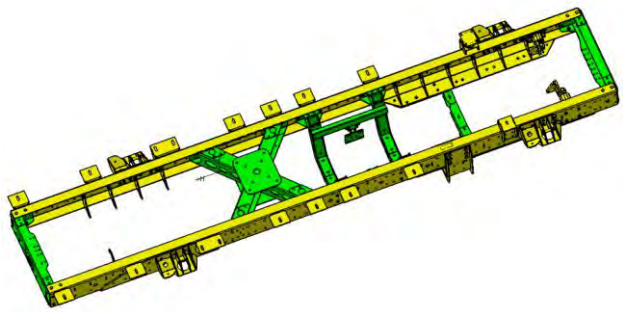


Figure 3 - Cross members

After fabrication, the cross member can be attached to the longitudinal members using either bolted joints or riveted joints, depending on manufacturing conditions and available equipment. Both joining methods can be used within the proposed concept. However, riveted joints are preferred where a permanent connection is required, since they reduce the possibility of loosening during vehicle operation. Bolted joints remain applicable in cases where easier disassembly or replacement of components is required.

In the area surrounding the transfer case, the cross members were additionally designed so that, with only minor modifications, they can also serve as supporting elements for a V-shaped blast protection structure, as shown in Figure 4. This solution provides the possibility of integrating an underbody mine protection system without requiring a complete redesign of the chassis structure, thereby preserving the modularity of the proposed concept.

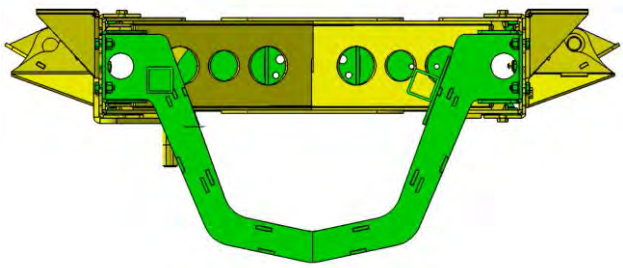


Figure 4 - Cross members around transfer case

The proposed cross member design contributes to improved torsional stiffness while maintaining manufacturing simplicity, assembly flexibility and the potential for future integration of additional structural and protective elements.

4.4 Adjustable wheelbase concept

One of the key features of the proposed chassis is the ability to modify the wheelbase without redesigning the primary load-bearing structure. In conventional chassis designs, changing the wheelbase generally requires the development of a new frame variant, resulting in additional manufacturing effort, dedicated tooling and increased logistical complexity.

In the proposed concept, the suspension subframes are mounted to a dedicated supporting structure that can be translated along the longitudinal members after loosening the corresponding fastening bolts. This arrangement enables straightforward adjustment of the wheelbase while preserving the same primary chassis structure and the majority of its structural components. Once the desired position is reached, the supporting structure, shown in Figure 5, is secured using the existing fastening points.

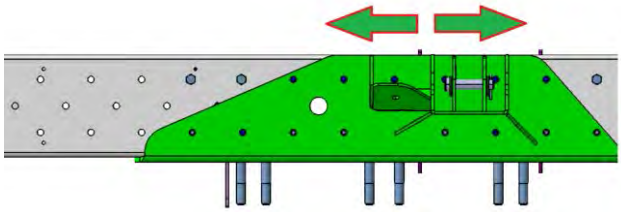


Figure 5 - Adjustable wheelbase concept

When a shorter wheelbase is required, the rear overhang can be reduced by shortening the longitudinal members to suit the new vehicle configuration. Conversely, if a longer wheelbase is required, the chassis can be extended without altering the basic structural concept. In practice, the maximum achievable chassis length is primarily limited by manufacturing constraints, such as the dimensions of the available sheet material and fabrication equipment.

Although changing the wheelbase requires corresponding modifications to the drive shaft, as well as portions of the electrical and pneumatic installations, these changes remain limited and straightforward. When the wheelbase is reduced, most of the existing components can be adapted to the new configuration without introducing additional parts. In cases where a longer wheelbase is required, only those components affected by the increased distance need to be replaced. Such components can be prepared as predefined conversion kits for each wheelbase configuration, enabling rapid vehicle reconfiguration without modifications to the primary chassis structure.

The proposed adjustable wheelbase concept significantly increases platform flexibility by enabling multiple vehicle configurations to be realized using the same basic chassis architecture. As a result, the number of dedicated chassis variants is reduced, while manufacturing, maintenance and logistical support are considerably simplified.

5. COMPARATIVE ANALYSIS

The design features presented in the previous section were developed with the objective of improving the manufacturability, modularity and operational flexibility of the proposed chassis concept. However, the practical value of these design solutions can only be assessed through a direct comparison with the existing chassis design. For this reason, a comparative analysis was performed to evaluate the influence of the proposed concept on manufacturing complexity, structural configuration and assembly.

Since the proposed vehicle platform incorporates both a redesigned chassis structure and an integrated suspension subframe, the comparison is carried out at two levels. First, the primary chassis structures are compared independently of the suspension subframes in order to isolate the effects of the structural redesign. Subsequently,

the complete chassis assemblies are compared to evaluate the overall benefits of the proposed vehicle platform from the standpoint of manufacturing, assembly and logistics.

5.1 Comparison of the primary chassis structure

The first stage of the comparative analysis focuses exclusively on the primary chassis structure, excluding the suspension subframes and associated mounting components. This approach enables an objective evaluation of the structural modifications introduced in the proposed chassis concept, independent of the influence of the selected suspension architecture.

A comparison of the existing and proposed chassis structures is presented in Figure 6. The comparison highlights the effects of the simplified structural design, interchangeable longitudinal members and standardized cross members on the overall chassis configuration.

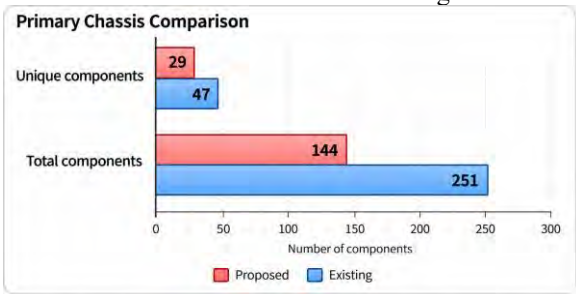


Figure 6 - Comparison of the primary chassis structure

The proposed chassis achieves a significant reduction in manufacturing complexity through a lower number of unique structural components and improved component standardization. In addition, the simplified structural layout facilitates manufacturing, assembly and maintenance while preserving the functionality required for different vehicle configurations.

5.2 Comparison of the complete chassis assembly

While the previous comparison focused exclusively on the primary chassis structure, the second stage of the analysis considers the complete chassis assembly, including the suspension subframes and all structural components directly associated with the chassis. This comparison provides a more realistic assessment of the manufacturing and assembly benefits achieved by the proposed concept.

As discussed in Section 4.2, the proposed chassis incorporates an integrated suspension subframe, which contributes to a further reduction in the number of structural components and assembly operations. To ensure an objective evaluation, the influence of this design decision was separated from the comparison of the primary chassis structure and is considered only in the complete chassis assembly, as shown in Figure 7.

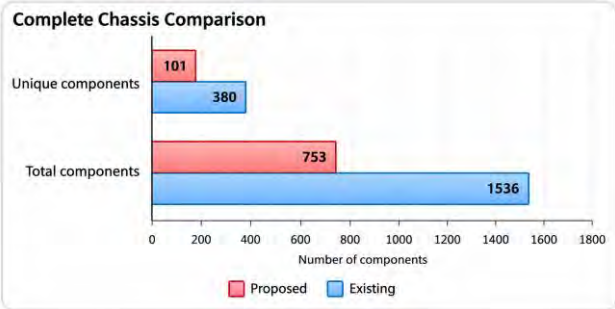


Figure 7 - Comparison of the complete chassis assembly

The complete chassis assembly of the existing vehicle consists of 1536 components, whereas the proposed concept reduces this number to 753. Components common to both vehicle concepts, such as the powertrain and other shared vehicle systems, were excluded from the comparison in order to evaluate only those components affected by the structural redesign. The reduction in the number of components simplifies assembly, inventory management, spare parts logistics and maintenance while supporting a more efficient manufacturing process.

5.3 Manufacturing and logistical benefits

The proposed chassis concept provides several manufacturing and logistical advantages compared to the existing design. The reduced number of structural components, together with the use of interchangeable longitudinal members and standardized cross members, simplifies production, assembly, inventory management and maintenance.

The adjustable wheelbase concept further increases platform flexibility by enabling different vehicle configurations to be realized using the same primary chassis structure. In cases where the wheelbase is increased, only the components directly affected by the modification need to be replaced using predefined conversion kits, while the primary chassis structure remains unchanged.

Overall, the proposed concept reduces manufacturing complexity, improves component standardization and simplifies logistical support while preserving the modularity required for modern military vehicle platforms. The reduction achieved by the proposed chassis concept is shown in Figure 8.

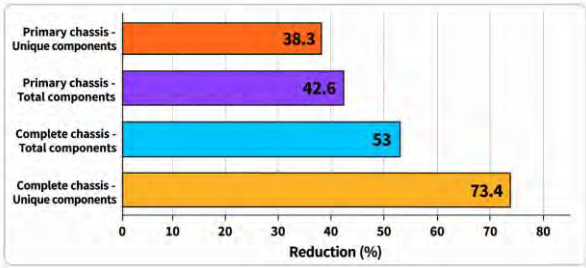


Figure 8 – The reduction achieved by the proposed chassis concept

6. CONCLUSION

This paper presented the development of a modular chassis concept intended to simplify the production and increase the operational flexibility of multi-purpose military vehicle platforms. The proposed design was developed with emphasis on simplified structural geometry, component standardization and the ability to adapt the wheelbase without redesigning the primary chassis structure.

Several design solutions were introduced to achieve these objectives, including interchangeable longitudinal members, standardized cross members and a movable suspension supporting structure enabling straightforward wheelbase adjustment. In addition to improving manufacturing flexibility, the proposed concept eliminates one of the common manufacturing errors associated with the production of left- and right-hand longitudinal members through the structural design itself.

A comparative analysis of the existing and proposed chassis concepts demonstrated a significant reduction in the number of structural components at both the primary chassis level and the complete chassis assembly level. The results indicate that the proposed concept can simplify manufacturing and assembly, reduce inventory requirements, facilitate maintenance and improve logistical support while preserving the modularity required for different vehicle configurations.

The proposed chassis concept represents a practical approach to the development of modular military vehicle platforms and provides a foundation for further structural verification, numerical analyses and experimental validation.

References

[1] DASCH,J.M., GORSICH,D.J.: *Survey of Modular Military Vehicles: Benefits and Burdens*, Defense Acquisition Research Journal, 23(1) (2016) 2-27.

[2] LAMPÓN,J.F., CABANELAS,P., GONZÁLEZ-BENITO,J.: *The impact of modular platforms on automobile manufacturing networks*, Production Planning & Control, 28(4) (2017) 335-348.

[3] IBRAHIM,F., AZALI,E.N.A., TAMILARASU,S., AB AZIZ,S.A., SARJO,N.A.: *Design and Optimisation of Military Vehicle Chassis: A Review*, Defence S&T Technical Bulletin, 18(2) (2025) 115-127.

[4] MALEN,D.E.: *Fundamentals of Automobile Body Structure Design*, SAE International, Warrendale, Pennsylvania, 2011.

[5] DIŽO,J., BLATNICKÝ,M., DROŹDZIEL,P., SEMENOV,S., MIKHAILOV,E., KURTULÍK,J.: *Strength analysis of an off-road lorry frame*, Scientific Journal of Silesian University of Technology. Series Transport, 110 (2021) 23-33.

[6] HODGES,S., OSTBERG,D.: *A Comparison of Structural Approaches for Combat Vehicles*, Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), Novi, Michigan, United States, August 13, 2014.

[7] THOTA,J.: Optimal Design of Vehicle with Internal Space Frame Structure Subjected to High Impact Load, PhD Dissertation, University of Nevada, Las Vegas, 2010.

[8] HU,S.J., ZHU,X., WANG,H., KOREN,Y.: *Product variety and manufacturing complexity in assembly systems and supply chains*, CIRP Annals - Manufacturing Technology, 57(1) (2008) 45-48.

[9] COMAN,M.N., RONTESCU,C., BOGATU,A.M., CICIC,D.T.: *Analysis of robotic welding possibilities of a car chassis assembly*, IOP Conference Series: Materials Science and Engineering, 1182 (2021) 012015.

[10] SVENSSON,L.-E., KARLSSON,L., SÖDER,R.: *Welding enabling light weight design of heavy vehicle chassis*, Science and Technology of Welding and Joining, 20(6) (2015) 473-482.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTERNAL BALLISTICS-BASED DESIGN AND STRUCTURAL ANALYSIS OF A 30×165 mm CANNON BARREL

MIHAILO BOJKOVIĆ
Military Technical Institute, Belgrade, bojkovicmih@gmail.com

VLADIMIR DUNIĆ
University of Kragujevac, Faculty of Engineering, , dunic@kg.ac.rs

VLADIMIR MILOVANOVIĆ
University of Kragujevac, Faculty of Engineering, , vladicka@kg.ac.rs

Abstract: *This paper deals with the analysis and optimization of parameters influencing the internal ballistics and barrel design of the 30×165 mm automatic cannon. Internal ballistic calculations were performed using a mathematical model based on the Drozdov method, with the aim of determining the influence of various parameters on projectile muzzle velocity and maximum powder gas pressure inside the barrel. The analysis was first carried out using reference internal ballistic parameters, after which selected parameters were varied within a limited range in order to examine their influence on system performance. By individual and combined variation of the parameters, their mutual dependence and impact on the operating characteristics of the weapon system were determined. The primary objective was to obtain higher projectile muzzle velocity while maintaining the maximum pressure within allowable limits. Based on the obtained optimal parameters, the barrel geometry was calculated. In the final stage of the research, structural verification of the designed barrel was performed using the finite element method in Simcenter Femap. The barrel model was subjected to loading conditions corresponding to the calculated internal ballistic forces, enabling analysis of stress distribution, deformation, and structural integrity under operating conditions.*

Keywords: *Internal ballistics, barrel design, Drozdov method, FEM analysis.*

1. INTRODUCTION

The design of artillery gun barrels requires the simultaneous consideration of internal ballistic performance and structural integrity. During firing, the barrel is subjected to rapidly varying pressure loads generated by the combustion of the propellant, producing complex stress and deformation states. Therefore, reliable barrel design requires both accurate prediction of the internal ballistic parameters and evaluation of the structural response under transient loading.

Internal ballistic calculations provide the pressure distribution, projectile velocity and projectile position throughout the firing process. These parameters can be used as input for finite element analysis, enabling a detailed assessment of the barrel stress and deformation during projectile travel. Such an approach allows the direct connection between the internal ballistic solution and the structural analysis of the weapon system.

In this study, an internal ballistic model of a 30×165 mm automatic cannon cartridge was used to determine the pressure distribution along the barrel. The obtained pressure curves were converted into a sequence of loading conditions and applied to an axisymmetric finite element model of the barrel developed in Simcenter Femap/Nastran. The structural response was evaluated by monitoring the maximum equivalent stress and total displacement throughout the firing cycle.

Furthermore, the structural response obtained for the optimized propellant configuration was compared with that of the initial configuration. The comparison enables the assessment of how the optimization of internal ballistic parameters influences the stress state and deformation of the barrel, providing a basis for further improvement of barrel design and material selection.

The proposed methodology provides an efficient procedure for transferring internal ballistic results into finite element load cases and can be applied during the structural design and verification of gun barrels.

2. INTERNAL BALLISTICS CALCULATION

The structural analysis presented in this paper is based on the results of an internal ballistic calculation performed for a 30×165 mm automatic cannon cartridge. The internal ballistic model was developed using the Drozdov method [1]. The calculation provides the temporal evolution of the chamber pressure, projectile velocity and projectile position, which are subsequently used as input data for the finite element analysis.

The initial calculation was performed using the nominal ballistic parameters of the selected ammunition. Input

Corresponding author: M. Bojković
E-mail: bojkovicmih@gmail.com

DOI: 10.5937/OTEH26038B

parameters for calulation are summarized in Table 1.

Table 1. Nominal input parameters

Label	Value	Unit
Powder chamber volume- W_0	0.000311	m
Barrel bore cross-sectional area- S_c	0.000738	m^2
Total projectile barrel travel- X_u	2.16	m
Forcing pressure- p_0	30e6	Pa
Fictitiousness coefficient- φ	1.169	/
Projectile mass- m	0.435	kg
Propellant charge mass- m_b	0.195	kg
Propellant density- ρ_b	1600	$\frac{kg}{m^3}$
Covolume of propellant gases-α	0.00103675	$\frac{m^3}{kg}$
Specific heat ratio- θ	0.269	/
Propellant grain thickness- r_0	0.000335	m
Shape coefficient- κ_1	1.24179	/
Shape coefficient- λ_1	-0.19471	/
Specific work of propellant gases- f_b	926712.491	$\frac{J}{kg}$
Propellant burning rate- u_{z0}	6.66e-10	$\frac{m}{sPa}$

The results of the IB calculations indicate a maximum chamber pressure of 313.69 MPa and a muzzle velocity of 1008.95 m/s.

Graphical results of IB calculation are shown in images below.

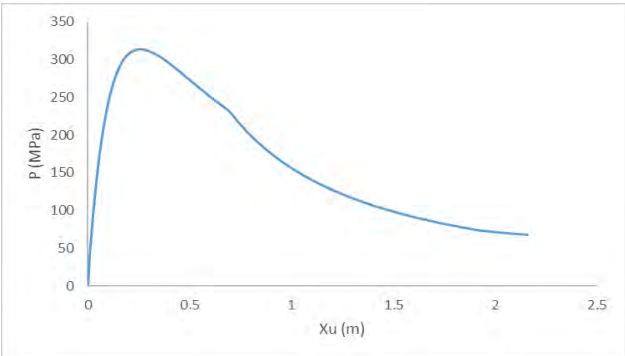


Figure 1. Pressure-travel curve.

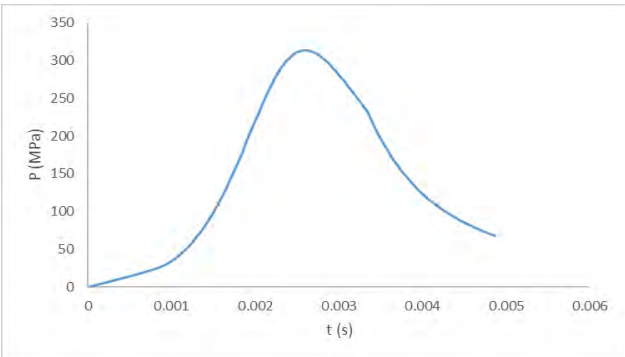


Figure 2. Pressure-time curve.

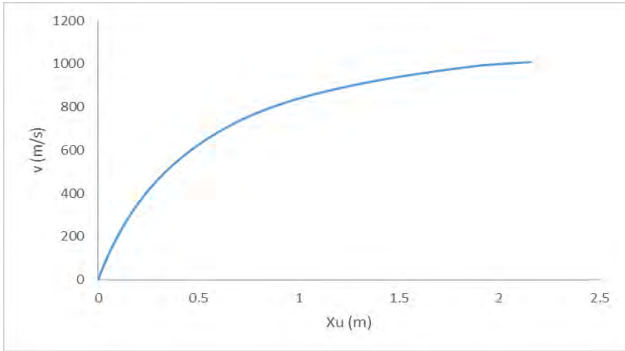


Figure 3. Velocity-travel curve.

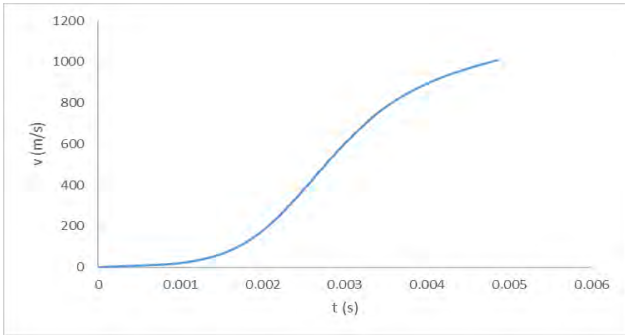


Figure 4. Velocity-time curve.

The initial ballistic solution served as the reference configuration for the subsequent optimization process.

2.1. Selection of the optimized IB configuration

The influence of the most significant internal ballistic parameters on the muzzle velocity and maximum chamber pressure was investigated by varying each parameter individually within predefined limits. [2] The objective of the analysis was to identify the parameters whose variation produces a favorable increase in projectile velocity without causing an excessive increase in the maximum pressure. [3]

Only the parameters that demonstrated a positive influence on the desired ballistic performance were selected for further optimization. Parameters with negligible or unfavorable influence were excluded from the final configuration.

The selected parameters together with their influence on the muzzle velocity and maximum pressure are presented in Table 2.

Table 2. Influence of changed parameters.

Parameter	Parameter change (%)	Velocity change (%)	Pressure change (%)
m	-5%	+1,64%	-5,03%
r_0	-5%	+1,72%	+10,92%
m_b	+2,5%	+1,85%	+6,37%
m_b	+5%	+4,63%	+16,72%
u_{z0}	+5%	+1,62%	+10,2%

α	+5%	+0,83%	+3,22%
f_b	+2,5%	+1,67%	+4,99%
f_b	+5%	+3,31%	+10,11%

Based on the observed trends, an optimized combination of internal ballistic parameters was defined. Newly selected parameter changes are: m-5%; m_b+2.5%; α+5%; f_b+2,5%. The selected parameter values represent a compromise between increased projectile velocity and acceptable structural loading of the barrel. This optimized configuration was adopted as the basis for the subsequent structural analysis. [4]

Graphical results of IB calculation are shown in images below.

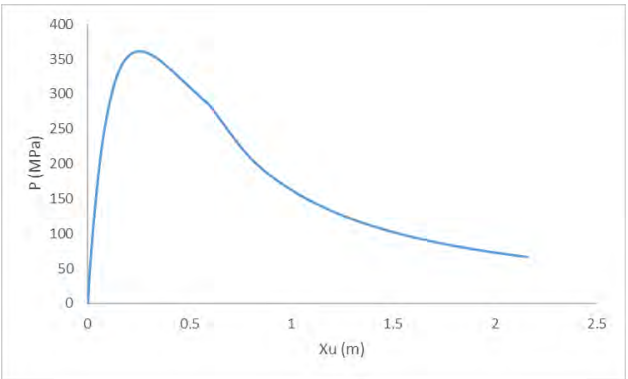


Figure 5. Pressure-travel curve.

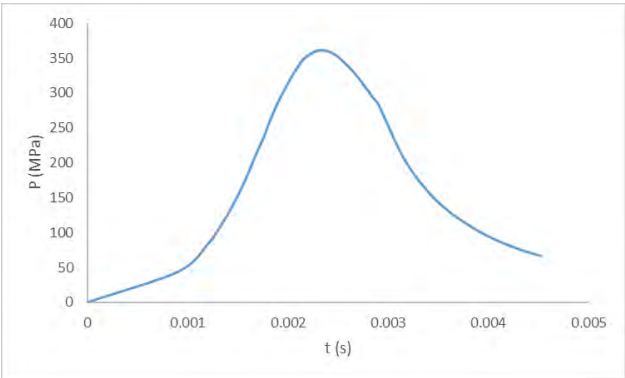


Figure 6. Pressure-time curve.

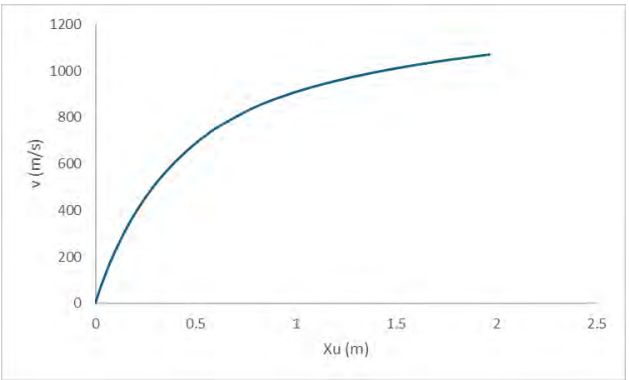


Figure 7. Velocity-travel curve.

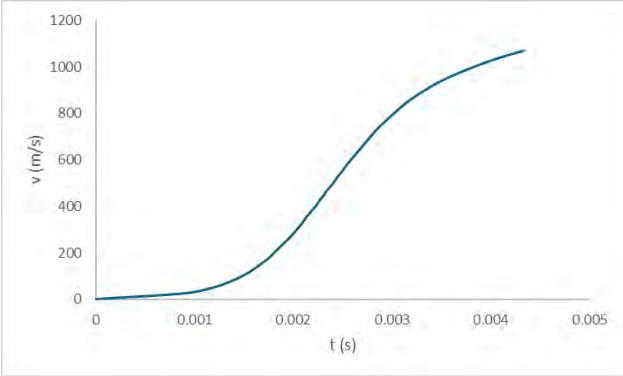


Figure 8. Velocity-time curve.

A comparison between the initial and optimized internal ballistic configurations is presented in Table 3.

Table 3. Comparison of the ballistic configurations.

	Maximum pressure (MPa)	Muzzle velocity (m/s)
Nominal conifig.	313.69	1008.95
Optimized config.	361.88	1090.49
Relative change	+15.36%	+8.08%

3. FINITE ELEMENT MODEL OF THE BARREL

The structural response of the barrel was evaluated using the finite element method based on the pressure distribution obtained from the internal ballistic analysis. The complete procedure included the development of the barrel geometry, generation of the finite element model, definition of material properties and boundary conditions, and application of the pressure loading corresponding to different projectile positions during the firing process.

3.1. Barrel geometry

The barrel geometry was generated using the KODCEV software package based on the optimized internal ballistic solution. The software calculates the required barrel profile from the specified ballistic loading conditions and material characteristics, providing the inner and outer barrel contours required for structural analysis. The barrel was dimensioned according to the Huber-Mises-Hencky (von Mises) criterion, where the elastic strength limit is defined by internal pressure at which the equivalent stress reaches the yield strength of the material. [5]

The calculated barrel profile was exported as a set of coordinate points describing the barrel cross-section. These coordinates were imported into the FEA environment, where the two-dimensional axisymmetric geometry of the barrel was reconstructed. Since the loading and geometry are rotationally symmetric, an axisymmetric representation was adopted in order to reduce the computational cost while preserving the accuracy of the structural analysis.

Due to space limitations, profiles of the chamber, the

forcing cone and the initial section of the projectile guide are represented in figure 9.

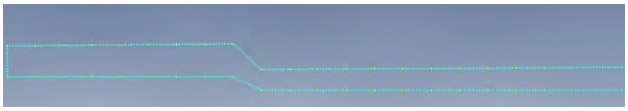


Figure 9. Profiles of the chamber, the forcing cone and the initial section of the projectile guide.

3.2 Finite element model

The finite element model was developed in Simcenter Femap**.** The reconstructed barrel profile was meshed using two-dimensional axisymmetric solid elements. The final mesh consisted of a combination of quadrilateral and triangular elements, with a characteristic element size of approximately 2 mm. This element size provided sufficient spatial resolution along the barrel profile while maintaining reasonable computational efficiency. The resulting finite element model contained a total of 5719 nodes and 4671 elements.

Linear elastic material properties corresponding to the selected barrel steel were assigned to the model. The material was assumed to be homogeneous and isotropic throughout the analysis.

The rear end of the barrel was constrained to represent its attachment to the breech mechanism. This boundary condition prevented rigid body motion while allowing the structural response of the barrel to be evaluated under the applied internal pressure loading.

The completed finite element model together with the applied boundary conditions is presented in Figure 10.

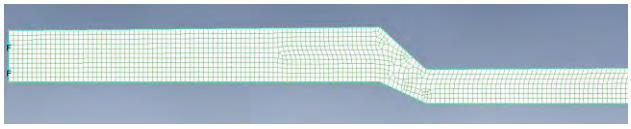


Figure 10. Finite element model with the applied boundary conditions.

3.3. Pressure loading

The pressure loading applied in the finite element model was derived directly from the optimized internal ballistic calculation. The pressure acting inside the barrel changes continuously as the projectile travels along the bore. Therefore, instead of applying a single uniform pressure, the loading was represented by a sequence of pressure distributions corresponding to successive projectile positions.

The pressure history obtained from the internal ballistic calculation was discretized into 40 loading steps, each representing a characteristic projectile position inside the barrel. For every loading step, the pressure was applied only to the portion of the barrel occupied by the propellant gases, while the remaining part of the barrel remained unloaded. This procedure reproduces the actual movement of the pressure front during the firing process.
To automate the preparation of the finite element model, custom BASIC scripts were developed within Simcenter Femap. The scripts automatically generated the pressure functions, created the corresponding Load Sets and assigned the pressure distributions to the appropriate barrel segments. The use of custom automation scripts enabled an efficient transfer of internal ballistic results into the finite element model, substantially reducing the time required for model preparation and minimizing the possibility of human error. [6]

Example of the sequence of pressure loading conditions applied during the analysis is schematically illustrated in Figure 11.

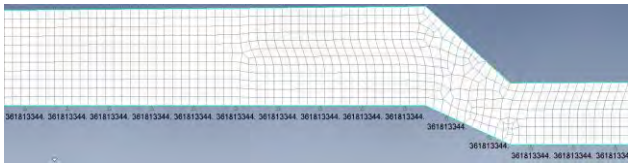


Figure 11. Example of the sequence of pressure loading conditions.

3.4. Numerical analysis

Following the definition of the finite element model and loading conditions, a MultiSet analysis was performed using Simcenter Nastran. Each loading step corresponded to one projectile position, allowing the evolution of the stress and deformation fields to be monitored throughout the firing process.

For every load case, the maximum equivalent (von Mises) stress and total displacement were extracted. These quantities were subsequently used to compare the structural response of the initial and optimized internal ballistic configurations. Also, an element and node ID of located maximum stress and displacement were also extracted.

The obtained results are presented and discussed in the following section.

4. RESULTS AND DISCUSSION

The structural response of the barrel was evaluated throughout the complete firing cycle using the sequential pressure loading obtained from the nominal and optimized internal ballistic calculation. For each loading step, the maximum equivalent (von Mises) stress and total displacement were extracted and analyzed as functions of both projectile position and time.

The variation of the maximum equivalent stress with projectile position and firing time is presented in Figures 12 and 13, respectively.

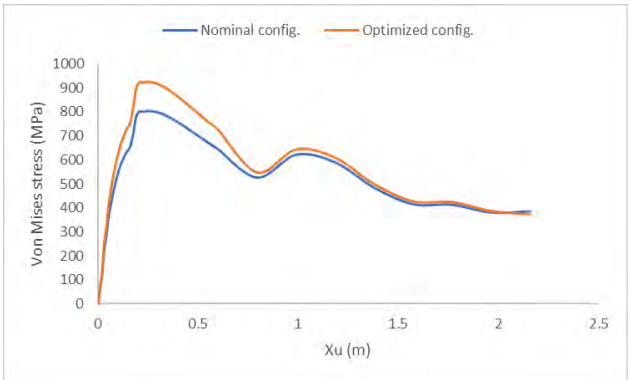


Figure 12. Von Mises stress-travel curve.

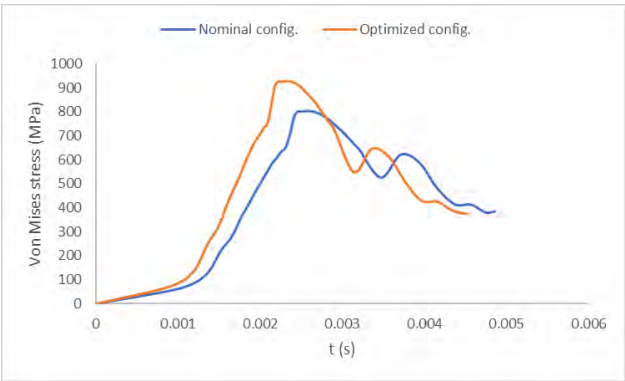


Figure 13. Von Mises stress-time curve.

Both figures demonstrate that the equivalent stress rapidly increases during the initial stage of projectile motion, reaches its maximum shortly after the pressure peak, and gradually decreases as the projectile approaches the muzzle. This behavior follows the evolution of the internal pressure generated during the firing process.

The variation of the total barrel displacement with projectile position and firing time is shown in Figures 14 and 15.

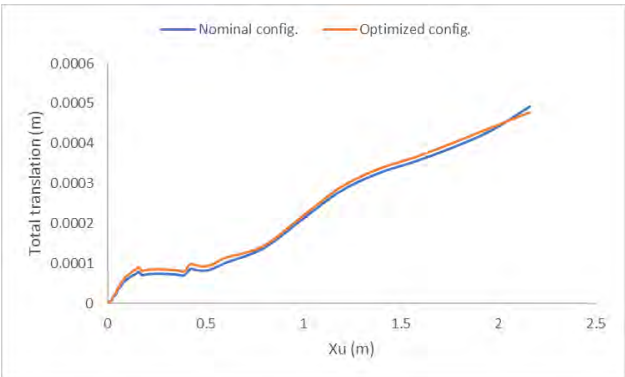


Figure 14. Total translation-travel curve.

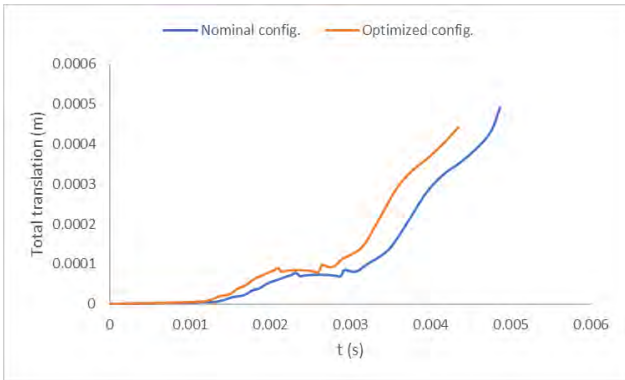


Figure 15. Total translation-time curve.

Unlike the equivalent stress, the total displacement exhibits a gradual increase during projectile travel, reaching its maximum near the muzzle. This behavior is expected since the loaded length of the barrel continuously increases as the projectile advances, resulting in progressively larger elastic deformation of the structure.

To evaluate the influence of the internal ballistic optimization, the maximum equivalent stresses obtained for the initial and optimized configurations were compared. The comparison is presented in table 4.

Table 4. Comparison of the maximum equivalent stress.

Configuration	Maximum equivalent stress (MPa)
Nominal	803.11
Optimized	926.75

The optimized internal ballistic configuration resulted in a increase of the maximum equivalent stress while simultaneously increasing the muzzle velocity. This indicates that the selected combination of ballistic parameters improves ballistic performance but decreases structural loading conditions of the barrel.

Although the calculated equivalent stresses reach relatively high values, the duration of the loading is extremely short. Therefore, the maximum equivalent stress was not used as the sole criterion for material selection based on the yield strength. Instead, the obtained stress distribution was primarily used to identify the most highly loaded regions of the barrel and to evaluate possible design improvements.

The numerical results indicates one characteristic region in which the maximum equivalent stresses repeatedly occurs during the firing cycle. The region is located near the transition between the chamber and the projectile guiding section. The repeated occurrence of stress maxima in this region suggests that further structural optimization could be achieved by modifying local geometric features, such as transition radii or wall thickness distribution. This region is shown in Figure 16.

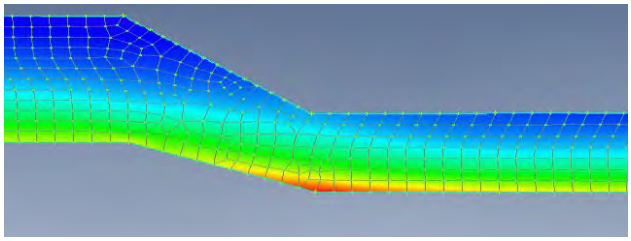


Figure 16. Repeated maximum stress occuring region.

From the manufacturing point of view, the obtained stress levels also indicate that additional strengthening methods may be considered in the most highly loaded portions of the barrel. Depending on the selected material and production technology, suitable solutions include local surface strengthening of the projectile guiding section, strengthening of a limited wall thickness, or the application of an autofrettage process in order to improve the fatigue resistance and operational life of the barrel.

The displacement results may also be useful during the design of the complete weapon system. Appropriate stiffening elements or barrel support arrangements may therefore be introduced to reduce deflection and improve the overall structural rigidity of the weapon.

Overall, the presented methodology demonstrates that coupling internal ballistic calculations with finite element analysis provides not only an assessment of barrel stresses and deformations but also valuable information for structural optimization and manufacturing process selection.

5. CONCLUSION

A methodology for the coupled internal ballistic and finite element analysis of a 30×165 mm automatic cannon barrel has been presented. The pressure distribution obtained from the internal ballistic calculation was transferred into a sequence of finite element loading conditions, enabling the structural response of the barrel to be evaluated throughout the firing cycle.

The parametric optimization of the internal ballistic model resulted in an increase in muzzle velocity, while the optimized pressure history provided realistic loading conditions for the finite element analysis.

The numerical analysis identified the evolution of equivalent stresses and barrel displacements and the most highly loaded regions. These results can be used to identify locations where additional strengthening methods, such as local surface treatment or autofrettage**,** may improve barrel service life.

The presented methodology demonstrates that coupling internal ballistic calculations with finite element analysis provides an efficient approach for the structural assessment and optimization of gun barrels, offering practical guidance for their design and further development.

References

[1] TANČIĆ, Lj.: Collection of internal ballistics tasks, Sector for education, training, scientific and publishing activities, Military technical academy of the army of Yugoslavia, Belgrade, 1999.

[2] KALEZIĆ M.: Design of artillery systems, Belgrade, 2010.

[3] TANČIĆ, Lj.: Internal ballistics design, textbook, Military Academy, Belgrade, 2014.

[4] MILOVANOVIĆ M. et al.: Design of mortar barrel 82 mm on basis of analysis of internal ballistics parameters, 11th International scientific conference on defensive technologies OTEH 2024, Tara, Serbia, 09.-11. October 2024. p.109 -113.

[5] RISTIĆ Z., JAKOVLJEVIĆ M.: Fundamentals of Armament, textbook, Military Academy, Belgrade, 2001.

[6] ŽIVKOVIĆ M.: Nonlinear structural analysis, Faculty of Mechanical Engineering, Kragujevac, 2005.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CONCEPTUAL ANALYSIS OF A RADIAL HYDRAULIC MOTOR DRIVE FOR HEAVY DUTY MILITARY VEHICLES

MIHAILO ĐUKOVIĆ
Faculty of Engineering University of Kragujevac, Kragujevac, mihailo.djukovic.02@gmail.com

SAŠA MILOJEVIĆ
Faculty of Engineering University of Kragujevac, Kragujevac, sasa.milojevic@kg.ac.rs

SLOBODAN SAVIĆ
Faculty of Engineering University of Kragujevac, Kragujevac, ssavic@kg.ac.rs

MILAN BUKVIĆ
The General Staff of the Serbian Armed Forces, Belgrade, milanbukvic76@gmail.com

HORIA ABAITANCEI
Institute of Automotive Engineering Transilvania University of Brasov, Brasov, h.abaitancei@unitbv.ro

BLAŽA STOJANOVIĆ
Faculty of Engineering University of Kragujevac, Kragujevac, blaza@kg.ac.rs

Abstract: *The increasing demand for highly mobile specialized vehicles has encouraged the development of alternative drive concepts capable of improving mobility, reliability, and operational flexibility. Radial hydraulic motors represent a promising solution due to their favorable torque characteristics and efficient operation under demanding conditions. This paper investigates the potential application of radial hydraulic motors in heavy wheeled vehicle drive systems. The proposed concept is analyzed from the perspectives of traction performance, power distribution, and operational efficiency. Special consideration is given to vehicle operation in challenging terrain conditions where hydraulic drive systems may provide advantages over conventional mechanical transmissions.*

Keywords: *radial hydraulic motor, hydraulic transmission, heavy vehicles, off-road mobility, vehicle drive systems.*

1. INTRODUCTION

Modern heavy-duty military vehicles require high mobility, reliability, and survivability under severe off-road operating conditions. Overcoming harsh terrains and maintaining combat readiness under high thermal loads necessitate advanced drivetrain configurations. Consequently, developing high-performance propulsion architectures remains a critical research focus in modern military engineering [1].

Conventional mechanical transmissions offer high efficiency but suffer from structural complexity, substantial weight, and rigid torque distribution in multi-axle platforms. These limitations drive the investigation of alternative layouts. In modern vehicle design, individual in-wheel electric drives significantly improve traction control and maneuverability through independent torque distribution. However, while electric hub drives suit light passenger platforms, heavy combat vehicles operating under severe impact and continuous thermal loads require extreme power densities, making hydrostatic radial piston drives a promising alternative for heavy-duty applications [2, 3].

Hydrostatic systems transmit substantial power within compact dimensions, optimizing vehicle packaging and off-road adaptability. Comparative evaluations of electric, hydraulic, and pneumatic rotary actuators indicate that while pneumatic systems offer excellent environmental resilience, their lower torque-to-weight ratios make them unsuitable for high-inertia military platforms [3]. Among hydraulic solutions, multi-stroke radial piston motors remain highly promising due to their High-Torque Low-Speed (HTLS) characteristics, making them ideal for direct wheel drives requiring high traction capability and precise torque control [4 - 9].

The objective of this paper is to investigate radial piston hydraulic motors in heavy multi-axle drivetrains using a multi-domain simulation model developed in *Simcenter Amesim*. The scientific contribution of this work is the development of a mathematical framework for quantifying internal volumetric leakage and thermal loading under continuous gradient climbing conditions, providing design guidelines for future heavy tactical vehicle applications.

Corresponding author: M. Đuković;
E-mail: mihailo.djukovic.02@gmail.com

DOI: 10.5937/OTEH26039D

2. RADIAL HYDRAULIC MOTOR PRINCIPLE

The Radial Piston Motor presents a crucial actuating component within hydraulic systems and represents a prominent variant of low-speed, high-torque motors (Figure 1.). It finds widespread application in sectors such as national defense, metallurgy, and aerospace. In order to preserve the environment, hydraulic actuators are increasingly expected to exhibit attributes such as low noise, minimal pollution, and smooth operation [8, 9].

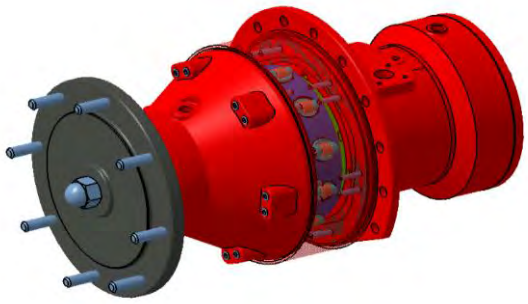


Figure 1. 3D model of Radial piston hydraulic motor

2.1. Construction and operating principle

A multi-stroke radial piston hydraulic motor features a compact and robust configuration designed to withstand high pressure spikes and harsh environmental conditions typical for heavy-duty military vehicles. It consists of the following key components (figure 2.) [10]:

- Cam-profile Stator: The stationary outer housing with an internal multi-lobe wavy profile. This component defines the stroke path and number of working cycles per revolution.
- Rotor (Cylinder Block): The rotating core containing radially arranged cylinder chambers. It is directly coupled to the wheel hub or output drive shaft via a heavy-duty spline.
- Piston Assemblies with Rollers: Reciprocating pistons housed within the rotor chambers. Each piston features a high-strength mechanical roller at its outer end, maintaining continuous contact with the stator profile to transmit the driving force.
- Flow Distribution Pair: The heavy-duty fluid commutation mechanism. It contains isolated internal passages that sequentially and accurately direct high-pressure oil to the cylinders and vent low-pressure spent oil back to the reservoir
- Output Drive Shaft: The heavy-duty shaft that extends outside the motor to drive the vehicle's wheel hub or sprocket.
- Motor Housing (Casing): The stationary outer housing with an internal multi-lobe wavy profile.

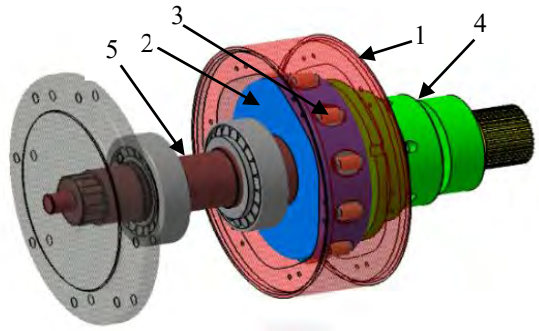


Figure 2. Key parts of radial hydraulic motor

The operational principle relies on converting fluid hydraulic energy into high mechanical output torque through cyclic reciprocating movement [10]:

- Inlet Phase (Working Stroke): Pressurized oil passes through the distribution pair (4) and enters the specific cylinder chamber. The high fluid pressure drives the piston (3) radially outward. As the piston roller forces its way down the inclined surface of the stator lobe (1), it yields a high tangential force, forcing the rotor (2) and the vehicle drive shaft (5) into rotation.
- Dead Centers: When the piston reaches its maximum outward displacement (Top Dead Center), the distribution slot isolates the chamber from the high-pressure line.
- Outlet Phase (Return Stroke): Guided by the continued rotation of the rotor, the descending stator profile mechanically pushes the piston radially inward. Concurrently, the distribution pair links the chamber to the low- pressure drainage network, discharging the oil to the reservoir.

The continuous repetition of these operating cycles generates a smooth rotational motion of the output shaft while maintaining high torque at low rotational speeds.

In the context of dynamic combat environments, understanding these micro-clearances is vital. The intense operating pressures within the piston-cylinder walls and across the sliding interface of the flow distribution pair create internal volumetric leakages. Mitigating these leakages is critical to preserving the volumetric efficiency and continuous traction of heavy- duty military drivetrains [3, 10, 11].

2.2. Main characteristics and comparison with conventional drivetrains

Radial piston hydraulic motors are inherently classified as HTLS displacement machines. Due to their multi-stroke architecture where each piston executes multiple working cycles per single rotor revolution, these units develop exceptionally high displacement volumes within a remarkably compact envelope [3]. This design delivers ripple-free, stable torque even at near-zero speeds, which represents a critical advantage for heavy military platforms navigating cross-country obstacles. Furthermore, their outstanding starting torque efficiency

ensures that fully loaded armored vehicles can initiate motion or resume climbing on steep, unstable inclines without any dangerous rollback [4, 6].

When integrated into heavy-duty tactical propulsion systems, radial piston configurations offer distinct structural and operational advantages over conventional mechanical and axial hydraulic alternatives. Standard mechanical drivetrains rely on complex networks of rigid drive shafts, multi-speed transfer cases, and heavy differentials. Implementing a radial hydrostatic drive eliminates these components entirely by utilizing an in-wheel hub-motor configuration. This architectural shift significantly reduces the structural weight of the platform, lowers its overall silhouette, and dramatically increases ground clearance to enhance off-road mobility [4, 6].

When compared to axial piston motors, which are fundamentally high-speed, low-torque devices, the radial alternative proves far more suitable for direct wheel drives. Driving a heavy armored vehicle wheel with an axial motor necessitates the integration of bulky planetary reduction gearboxes to multiply torque. Conversely, radial piston motors drive the wheel hub directly without mechanical gear reduction. This reduction in the total number of moving parts lowers mechanical friction losses, simplifies maintenance, and minimizes the vehicle's vulnerability to ballistic damage in combat environments [4, 6].

3. APPLICATION OF RADIAL HYDRAULIC MOTORS IN HEAVY MILITARY VEHICLES

To meet the extreme mobility demands of a heavy 6x6 tactical vehicle, a closed-loop hydrostatic drivetrain architecture is proposed. This system completely replaces the traditional mechanical gearbox, drive shafts, and differentials. The operational concept and fluid routing are illustrated in the system block diagram (Figure 3) [7, 12].

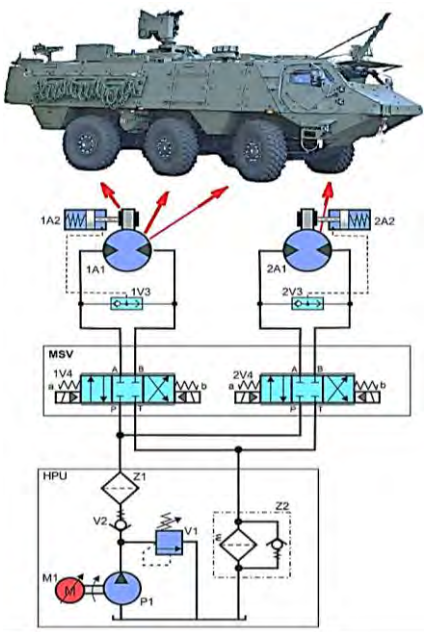


Figure 3. – Schematic diagram of conventional hydraulic propulsion system of the military vehicle

The schematic configuration of the proposed multi-wheeled 6x6 hydrostatic drive system consists of three main sub-assemblies: the Hydraulic Power Unit (HPU), the Main Valve Block (MSV), and the dynamic propulsion-braking hub stations. The fluid power is generated by the engine-driven pump P1, which is protected against overpressure transients by the safety relief valve V1. High-pressure filtration is handled via filter Z1, while return-line fluid conditioning and thermal management are achieved through the bypass-equipped radiator block Z2. Fluid flow rate and direction are independently regulated by two proportional directional valves, 1V4 and 2V4, embedded in the MSV block. Valve 1V4 directly controls the fluid delivery to the left-side radial piston hydraulic motor 1A1, which simultaneously drives all three left-side wheels through parallel porting lines. Conversely, valve 2V4 governs the right-side radial piston motor 2A1, enabling asymmetric flow adjustments for precise skid-steering maneuvers under tactical off-road scenarios [7].

3.1. Mathematical model

To evaluate the conceptual performance of a high-torque radial piston hydraulic motor for heavy-duty military operations, a comprehensive mathematical model is established. This model defines the relationship between hydraulic input energy, internal mechanical and volumetric losses, and the dynamic output performance required to propel a heavy military vehicle under severe off-road conditions.

The rotational speed of the radial hydraulic motor is directly governed by the volumetric flow rate of the hydraulic fluid supplied to its cylinders.
The theoretical flow rate Q_{th} [m³ s^{-1}], which assumes zero leakage and ideal geometry, is defined as [10, 13]:

$$Q_{th} = V_d \cdot n \tag{1}$$

Where:

- V_d is the displacement of the radial motor [m³ rev^{-1}];
- n is the rotational speed of the motor shaft [rps].

In heavy-duty military applications, high operating pressures lead to internal fluid bypass across the piston-cylinder clearances and the distribution valve. Therefore, the actual flow rate Q_{act} required by the motor is expressed as [13]:

$$Q_{act} = Q_{th} + Q_{leak} \tag{2}$$

Where Q_{leak} represents the internal volumetric leakage flow rate [m³ s^{-1}].

Radial piston motors are categorized as Low-Speed High-Torque (LSHT) drives, making them ideal for overcoming steep slopes and rough terrain.

The ideal or theoretical torque T_{th} (Nm) generated by the motor depends strictly on its geometric displacement and the differential pressure [10, 13]:

$$T_{th} = \frac{V_d \Delta p}{2\pi} \quad (3)$$

Where:

$$\Delta p = p_{in} - p_{out} \quad (4)$$

is the pressure drop across the motor (Pa).

The actual torque T_{act}delivered to the vehicle's wheel or sprocket is lower than the theoretical value due to mechanical friction between moving components (pistons, rollers, and cam lobes). It is calculated as [13]:

$$T_{act} = T_{th} - T_{loss} = T_{th}\eta_{hm} \quad (5)$$

Where:

- T_{loss} is the total friction torque loss (Nm);
- η_{hm} is the hydro-mechanical efficiency of the motor.

The conversion of hydraulic power into mechanical power determines the overall tractive capability of the vehicle.

The hydraulic input power P_{hyd} (W) delivered by the system pump is given by [13]:

$$P_{hyd} = Q_{act}\,\Delta p \quad (6)$$

The effective mechanical power P_{mech} (W) available at the output shaft to drive the vehicle is defined as [13]:

$$P_{mech} = T_{act}\omega = T_{act}\frac{2\pi n_{rpm}}{60} \quad (7)$$

Where:

- ω is the angular velocity of the shaft (rad/s);
- n_{rpm} is the shaft speed in revolutions per minute (rpm).

Military vehicles operate under extreme temperature fluctuations (from −30 °C to +50 °C), which drastically alters oil viscosity and loss characteristics.

The internal leakage flow Q_{leak} can be mathematically modeled using the Wilson-based hydraulic loss approach [13]:

$$Q_{leak} = \frac{C_s V_d \Delta p}{\mu} \quad (8)$$

Where:

- C_s is the dimensionless laminar leakage coefficient of the radial motor;
- μ is the dynamic viscosity of the hydraulic fluid (Pa·s).

The volumetric efficiency η_v is then formulated as [13]:

$$\eta_v = \frac{Q_{th}}{Q_{act}} = 1 - \frac{Q_{leak}}{Q_{act}} \quad (9)$$

The torque loss T_{loss} comprises both Coulomb (constant) friction and viscous fluid shear friction within the motor housing [13]:

$$T_{loss} = (C_f V_d \Delta p) + (C_v V_d \mu\omega) \quad (10)$$

Where:

- C_f is the mechanical friction coefficient;
- C_v is the viscous drag coefficient.

Consequently, the total efficiency of the hydraulic drive η_t is expressed as [13]:

$$\eta_t = \eta_v \eta_{hm} = \frac{P_{mech}}{P_{hyd}} \quad (11)$$

4. SIMULATION MODEL AND DISCUSSION

To evaluate the hydrostatic powertrain concept for the 6x6 heavy-duty tactical vehicle, a comprehensive multi-domain simulation model was established within the *Simcenter Amesim* environment The interconnected structural architecture of this simulation framework is schematically detailed in Figure 4.

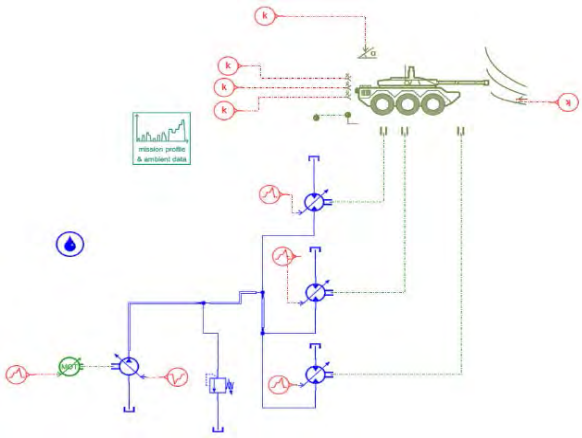


Figure 4. 6x6 concept combat vehicle drive diagram

The model couples the high-pressure fluid power network directly with the longitudinal vehicle dynamics of the 3-axle platform. The hydraulic propulsion layer consists of the primary pump unit supplying pressurized fluid to three parallel-connected in-wheel radial piston hydraulic motors, which model one complete side of the tactical vehicle. The advanced nature of this multi-domain model lies in its capacity to incorporate dynamic external environment boundaries (denoted by coefficient blocks k), which actively process terrain inclinations, aerodynamic drag, and variable tire-soil rolling resistance. The baseline parameters used in this simulation framework are summarized in Table 1 [12, 14, 15]

Table 1. Operational parameters of the 6x6 hydrostatic powertrain model

Quantity	Value
Total vehicle mass	25000 kg
Tire Specification	14R20
Prime Mover Speed	2000 rpm
Pump Displacement	600 [cm^3 rev^{-1}]
In-Wheel Motors Displacement	975 [cm^3 rev^{-1}]
Rolling resistance	0.02
Fluid density	850 [kg m^{-3}]
Dynamic viscosity	51 [cP]
Road slope	0 %
Wind velocity	10 [m s^{-1}]

The dynamic velocity response of the 25-tonne combat vehicle operating under a 10 [m s^{-1}] headwind profile is illustrated in Figure 5.

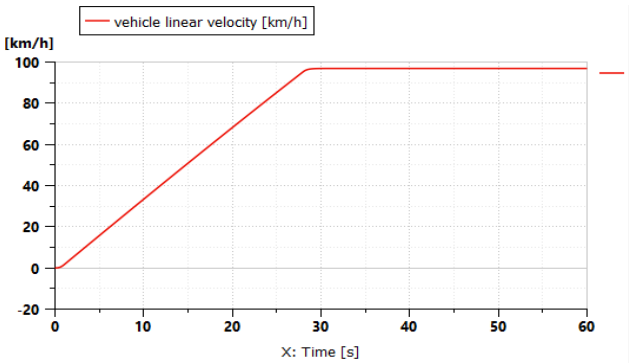


Figure 5. Diagram of vehicle velocity in 60s operational cycle

The simulated transient kinematic profile comprises three operational phases:

- System Activation (0–1 s): A 1-second response lag occurs due to the engine ramp-up to 2000 (rpm) and the pilot pressure buildup required to disengage the Spring-Applied Hydraulic-Released (SAHR) hub brakes.
- Linear Acceleration (1–22 s): Following brake release, the large 975 [cm^3 rev^{-1}] displacement of the in-wheel motors generates high torque, overcoming the static inertia of the 25000 (kg) chassis.
- Steady-State Dash (22–60 s): The vehicle reaches a maximum velocity of 97 km/h, establishing full hydrostatic equilibrium where wheel torque matches aerodynamic drag and rolling resistance.

The dynamic torque requirement at the wheel hub (Port 3) throughout the 60-second operational cycle is quantified in Figure 6.

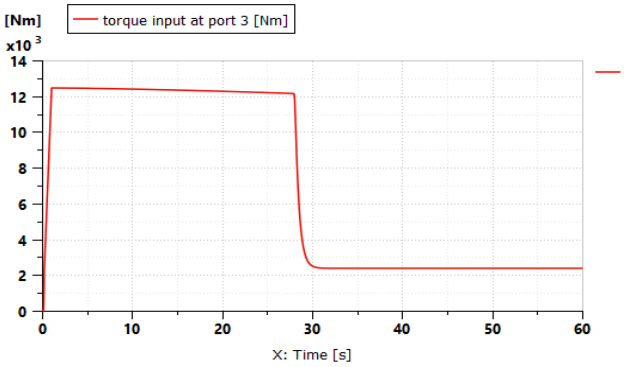


Figure 6. Simulated torque input profile at the axle

The torque curve validates the powertrain's capacity to sustain heavy-duty tactical movement:

- Inertial Launch (0–1 s): Drive torque peaks instantly at 12500 (Nm) to overcome the static inertia of the 25000 (kg) chassis and terrain resistance.
- Sustained Acceleration (1–28 s): Torque remains at an elevated plateau, shifting from 12500 (Nm) to 12200 (Nm), confirming maximum system pressure drop across the in-wheel motors.
- Steady-State Relaxation (28–60 s): At a cruise velocity of 97 km/h, torque drops rapidly between 28–30 s, settling at a 2400 (Nm) baseline to counteract rolling resistance and the 10 [m s^{-1}] wind drag.

To assess operational limitations under severe conditions, a separate simulation evaluates vehicle traction on a steep 15° road gradient (26.8% slope). To ensure a rigorous comparison, all hardware and fluid parameters remain identical to Table 1. This analysis focuses on how continuous gravitational loading alters drivetrain kinematics, pressure profiles, and volumetric leakage. Results are presented in Figures 7 and 8.

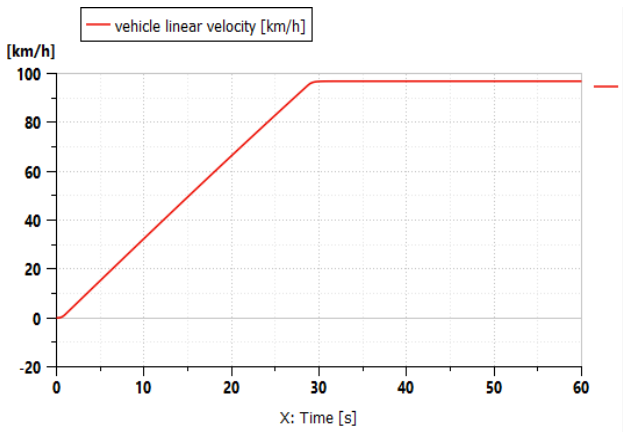


Figure 7. Diagram of vehicle velocity under the slope of 15 °

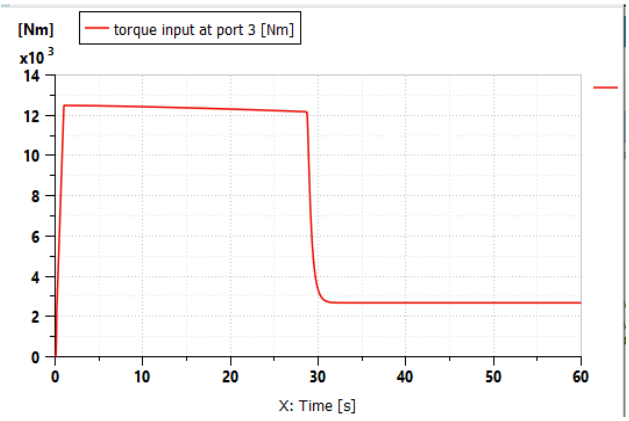


Figure 8. Diagram of simulated axle torque under the slope of 15 °

Although the drivetrain achieves 97 km/h on the 15° slope, grade resistance extends the acceleration phase and shifts hydrostatic equilibrium from 22 s to 30 s. This prolonged peak-pressure window forces fluid through the micro-clearances of the flow distribution pair and piston-cylinder walls, driving the volumetric leakage Q_{leak} to a higher plateau than on flat terrain. While the fluid dynamic viscosity of 51 [cP] at 40 °C prevents a total efficiency collapse, continuous high-load shearing generates severe thermal loads. This demonstrates that for multi-axle armored platforms executing tactical maneuvers on steep inclines, the flow distribution pair clearances must be precisely optimized to mitigate pressure-induced volumetric losses and prevent thermal thinning.

From a fluid power perspective, the operational discrepancy between flat terrain and the 15° slope highlights the volumetric allocation dynamics within the hydrostatic loop. On flat terrain, the main hydraulic pump's delivery flow rate rapidly satisfies the cumulative volumetric demand of the three parallel-connected wheel motors, reaching full flow equilibrium at 22 s. Conversely, during the steep climb, sustained gravitational resistance forces the wheel motors to operate under peak pressure drops to maintain the necessary torque output. This critical shift alters the flow rate distribution: while the main pump operates at a regulated displacement providing a baseline input flow, the high-pressure differential across the wheel motors drives fluid through internal clearances. Consequently, the individual flow rate utilized for effective mechanical work decreases due to the elevated volumetric leakage Q_{leak}. This behavior indicates that the efficiency bottleneck under severe thermal and mechanical loads is not the raw volumetric capacity of the main pump, but rather the localized leakage control within the flow distribution pairs of the in-wheel radial piston units.

5. CONCLUSION

This paper presented a conceptual analysis of an in-wheel hydrostatic drivetrain driven by multi-stroke radial piston motors for a 25-tonne 6×6 combat vehicle. The multi-domain simulation demonstrated that the proposed drive can provide sufficient tractive effort, achieving a maximum velocity of 97 km/h. Comparative analysis revealed that a steep 15-degree road inclination proportionally increases the system response time and shifts the stabilization threshold from the 22nd to the 30th second. This prolonged operation under maximum pressure conditions increases the significance of internal volumetric leakage within the flow distribution pairs. Although the dynamic fluid viscosity of 51 cP at 40 °C prevents a complete efficiency degradation, the sustained high-pressure gradient during severe slope operation generates considerable thermal loads. These findings highlight that precise CFD analysis and experimental optimization of flow distribution pair clearances are essential for improving the efficiency and reliability of advanced tactical wheeled platforms. Future work will focus on structural optimization of these critical micro-clearances and subsequent experimental validation using a dedicated hydraulic test rig.

References

[1] NUNEZ,P., JONES,R., SHOOP,S.: High-Fidelity Ground Platform and Terrain Mechanics Modeling for Military Applications, Technical Report, Defense Technical Information Center, Accession Number: ADA432776, (2004)

[2] POPOVICI, V., GRIGORAS, I. R., RAKOSI, E., ULIAN, T. M., MANOLACHE, G., POPA, A. G., PETROVAN, S.: A New Approach to In-Wheel Motor Solutions for Electric Vehicles, World Electric Vehicle Journal, 17(2) (2026) 87

[3] TANAKA,K., OKAZAKI,T., URABE,T.: Comparative Study on Dynamic Characteristics of Hydraulic, Pneumatic and Electric Motors, Proceedings of the ASME/BATH 2013 Symposium on Fluid Power and Motion Control, (2013) 10.1115/FPMC2013-4459

[4] ABAITANCEI,H., RADU,S., MUNTEAN.B., IRMIA,C.: Multi-Domain and Durability Analysis of a Hybrid Hydraulic Vehicle, (2016) 823:217-222

[5] LI,C., JIANG,T., LIU,C., XU,H., SHI,G.: Investigation of the leakage in the flow distribution pair of radial piston hydraulic motors through CFD analysis and experiments, Flow Measurement and Instrumentation, 96 (2024) 102555.

[6] MANNESMANN REXROTH: Hydrostatic Transmissions for Wheeled and Tracked Vehicles, Technical Handbook, Rexroth Hydraulics, (1998).

[7] DINDORF,R.: Functional Safety of the Hydraulic Drive Control System of a Tracked Undercarriage, The Archives of Automotive Engineering - Archiwum Motoryzacji, 103(1) (2024) pp. 21-37

[8] BAŠTA,T.M.: Mechanical hydraulics, Faculty of mechanics, Belgrade, 1986.

[9] SAVIĆ,V,: Oil hydraulics, hydraulic components and systems, Zenica, 1986.

[10] GIAVARINI,F.: Analysis and Simulation of a Radial Piston Hydraulic Motor, Master's degree thesis, Politecnico di Torino, (2022).

[11] MERITT H.E.: Hydraulic Control Systems, John Wiley & Sons, (1967).

[12] PATRIA LAND SYSTEMS: Patria 6x6 - Modern Protection and Extreme Mobility, Technical Specification Brochure, Oy Patria AB, Finland, (2023)

[13] RADOSAVLJEVIĆ,M., PETROVIĆ,R., ŽEŽELJ,B., BELOVIĆ,D., CVEJIĆ,S.: Mathematical Modeling and Experimental Tests of Working Parameters of Piston-Radial Pumps in Hydraulic Systems of Mobile Machines, Applied Sciences, 14(21) (2024) p. 9979.

[14] WONG, J. Y.: Theory of Ground Vehicles, 4th Edition, John Wiley & Sons, New York, (2008)

[15] SHELL INTERNATIONAL PETROLEUM: Shell Tellus S2 MX 46 - Industrial Hydraulic Fluid Technical Data Sheet, Shell Lubricants, Version 2.4, (2021)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PREDICTION ALGORITHM FOR THE IMPACT POINT PROJECTILE TRAJECTORY USED FOR DETERMINING CORRECTIVE SHOOTING PROCEDURE ELEMENTS

DRAGOSLAV R. PETROVIĆ
Retired Colonel, former chief of Ballistic department MIT,Belgrade, Jurija Gagarina 146/50, 11070 N. Beograd,
drdrpetrovic@gmail.com

Abstract: *The lack (frequently) of meteo-ballistic perturbations forces the operators of Artillery fire (cannon, howitzer) to use the so-called Corrective shooting procedure. It involves firing the first projectile, observing the place of its impact in relation to the target, and correcting the fire elements (aiming) for the second projectile. The minimum time of this process is as long as the flight of the projectile, increased by time of observed data transfer to fire position. Multiple reduction of this time can be achieved by knowing the position data of the first projectile on the ascending part of its trajectory. An algorithm for predicting impact (terminal) point of artillery projectile trajectory is given. It is based on trajectory parameters deviation related to Nominal trajectory, thus predicting range deviations without necessity real range observation after final flight first projectile.*

Key words: *Software, Ballistic Module extension, Artillery shooting*

1. INTRODUCTION

The paper presents method and the procedure for predicting the location of the projectile's fall related to the distant target from a gun. It is based on the measurement of projectile position in space at the initial part of the trajectory, before it reaches the height of the vertex Y_s – *Figure 1.*

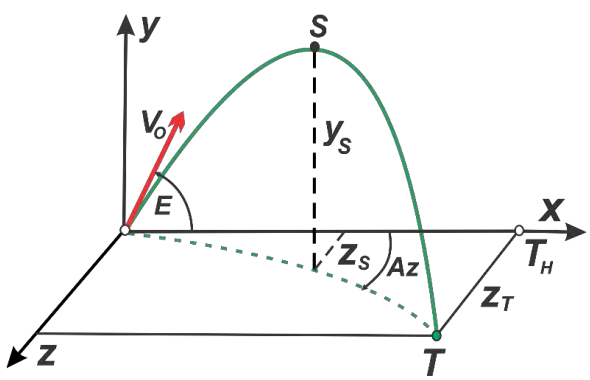


Figure 1. Projectile trajectory in space

For the desired distance of the gun to the target, according to the Ballistic Module -***BM*** -$[1,2,3]$, can be calculated required fire elements (***fe***) - elevation angle E in the vertical plane (x,y) and direction angle A_Z in the horizontal plane (x,z) -*Figure 1*. The practice in artillery units is to fire one projectile, observe the location of its fall and correct the initial ***fe***. Since the flight time of a long-range projectile is long by today's standards, and the problem is the observation and transmission of observed fall values, this method seems outdated and difficult to apply (deadly dangerous!). The predicted distance serves for correction initial (E , A_Z) as if it had been observed.

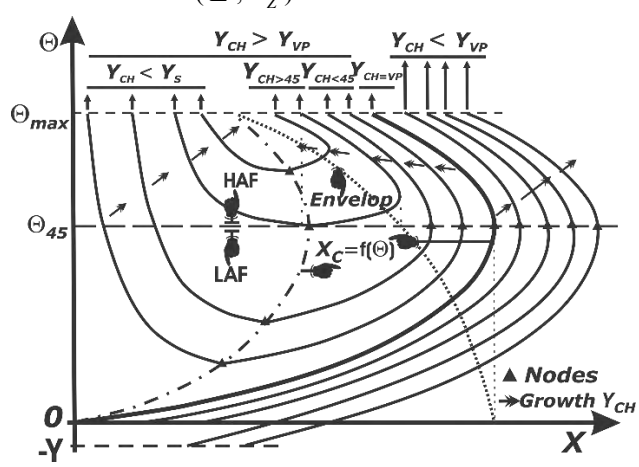


Figure 2. Dependance distance from elevation

For different relations between the altitudes of the gun Y_{VP} and the target Y_{CH} , relations $X, E(\theta)$ are given in the graphs, *Figure 2*. Elements (E , A_Z) can in principle be determined for meteo-ballistic conditions (***mb***), normal - nominal or actual. In Firing Tables - ***FT***

Corresponding author: D. Petrović;
E-mail: drdrpetrovic@gmail.com

DOI: 10.5937/OTEH26040P

are given trajectory parameters that have been designed for nominal ***mb***. In field conditions, it is difficult (nearly impossible) to find ***fe*** for current ***mb*** that can be used in ***BM***. This paper deals with overcoming the difficulty of this type, which enables calculation the prediction of the location where the projectile will fall related to the gun, without need to be observed fall and transferred to the gun position. Physical projectile is replaced by a mathematical model (***mm***). In ***mm*** are incorporated forces and moments acting on the projectile in flight.
The model is actually a system of differential equations -***de***, whose numerical solution gives parameters the projectile trajectory. With the replacement of large analog computers with digital ones, the era of digital miniaturization began, which is ongoing.

In this paper, it is assumed that complex mathematical operations can be performed on the projectile itself (hardware located in the body of the fuze on the projectile). For the numerical solution of ***de***, in addition to forces and moments, initial conditions and parameters of the environment - the atmosphere in which the projectile is moving - are required.

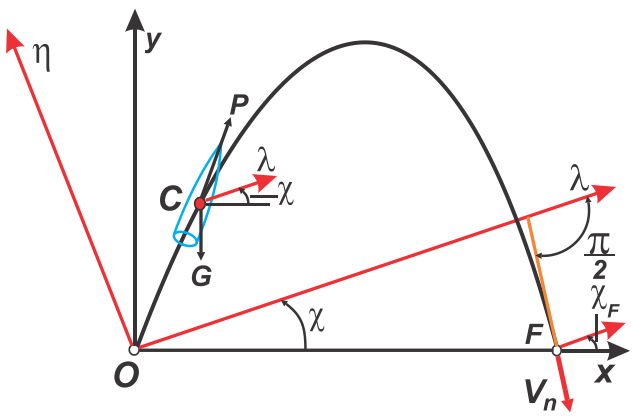


Figure 3. Inclined coordinate system λ, η

The initial conditions are called ballistic parameters, and the parameters of the atmospheric characteristics are called meteorological parameters, all together - ***mb***. When it comes to a ballistic rocket, a procedure similar to the prediction proposed here is carried out on the initial part of the trajectory, *called introduction time into the compensation integrator* -$[4,5,6]$. This procedure is based on the continuous minimization of the introduced ballistic function. For this reason, an inclined coordinate system λ, η of the longitudinal axis is used -*Figure 3*.

The angle of this axis χ is equal to the $tg(\chi_F)$ at the point of impact. Trajectory elements are calculated as if the rocket flight via atmosphere on the passive part of the trajectory is in vacuum. After ending minimization of the ballistic function, the rocket motor is turned off.

2. PROJECTILE DISTANCE PREDICTION METHOD

The flight of the projectile ends by falling in the surroundings of the target. It is required projectile to hit the target, which rarely happens because each hit is a random variable when it comes to an unguided projectile in consideration. The cause is the deviation of the actual trajectory from the nominal one – *Figure 4*, and the reason is the factors called disturbances, which are different from the nominal ones, on the basis of which the elements of the trajectory (E, A_Z) were calculated.

The idea on which the method of predicting the distance until the projectile falls in target surroundings the is based to measuring the parameters of the current trajectory. How to use obtained results and avoid certain difficulties and which mathematical operations should be performed, follows

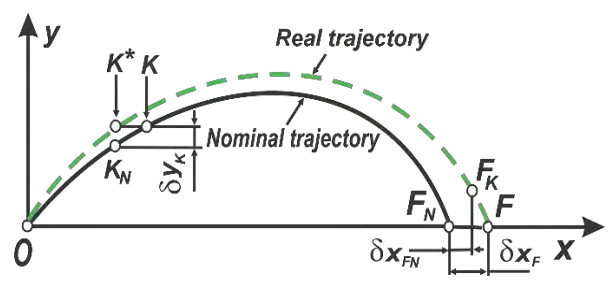


Figure 4. Nominal and disturbed trajectory

Let nominal and disturbed trajectory be described by the following differential equations:

$$\frac{dy_i^N}{dt} = f_i^N\left(y_i^N, z_j^N\right) \text{.....(1)}$$

$$\frac{dy_i^N}{dt} = f_i^N\left(y_i^N, z_j^N\right) \text{.....(2)}$$

where values y_i are the flight parameters, and z_j $(i = 1,...,n;\ j = 1,...,m)$ are the forces and moments acting on the projectile, as well as ***mb*** are the conditions in which the flight takes place. Let the solution of equation (1) $y_i^N = y_i^N(t)$ be obtained for the initial conditions y_i^0 given in point O- *Figure 3*. Let's assume that the real values (marked with *), y_i^* called perturbed, are in the form:

$$y_i^* = y_i^N + \Delta y_i;\ \ i = 1,...,n$$

$$z_j^* = z_j^N + \Delta z_j;\ \ j = 1,...,m \ \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

With (3) entered in (1,2) we get

$$\frac{d\Delta y_i}{dt} = f_i^*\left(y_i^N + \Delta y_i, z_j^N + \Delta z_j\right) -$$

$$-f_i^N\left(y_i^N, z_j^N\right) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

If the above equations are expanded into a Taylor series, we get

$$\frac{d\Delta y_i}{dt} = \sum_{i=1}^{n} p_{i,j}(t)\Delta y_i + \sum_{j=1}^{m} \pi_{i,j}\Delta z_j +$$

$$+f_i\ (t) + R_i\left(t, \Delta y_i, \Delta z_i\right) 1,2,3\ldots\ldots\ldots\ldots\ldots(5)$$

In (5), the sum R_i are terms which order is greater than the first, and

$$p_{i,j} = \frac{\partial f_i^*}{\partial y_j}\Big|\Delta y_j, \Delta z_j = 0$$

$$\pi_{i,j} = \frac{\partial f_i^*}{\partial z_j}\Big|\Delta y_j, \Delta z_j = 0 \qquad \ldots\ldots\ldots\ldots(6)$$

$$\Delta f_i = f_i^*\left(y_i^N, z_j^N\right) - f_i^N\left(y_i^N, z_j^N\right)$$

By discarding R_i from (5), the equation is obtained

$$\frac{d\Delta y_i}{dt} = \sum_{j=1}^{m} p_{i,j}(t)\Delta y_j + \sum_{j=1}^{m} \pi_{i,j}\Delta z_j +$$

$$+\Delta f_i\ (t)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

called the system equation of disturbed motion of the first approximation.

This system is (still!) used in consideration of the stability of the general movement with the condition that the disturbances are of small in size, which change during the flight, and in the case of a stable system tend to be zero. If we look at *Figure 4*, can be seen that at the altitude $Y = Y_K$ of the trajectory, the horizontal distance is equal to $K^* - K$. It grows during movement, so that it becomes the largest at the zero horizon $Y = Y_F = 0$, $Y = Y_F = 0$ whose knowledge is of key interest for us.

If the solutions ***de*** (7) were found numerically up to the time t_{F_N} obtained in the nominal ***mb***, the values $\delta x_K, \delta y_K$ would be obtained (*Figure 4*). By connecting the points F_K, F, a straight line is obtained $F_K - F$ whose angle is θ_F. Change of distance in the horizon with correction due to the altitude δy_{F_K} is:

$$\delta x_F = \delta x_{F_K} + ctg\,|\,\theta_F\,|\,\delta y_{F_K} \ \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

The presented procedure is of optional use because we are able to find distances $K^* - K$ on the basis of measuring the position of the projectile, but we have no way to determine the causes that led to these differences in order to calculate according to ***de*** the parameters of the disturbed trajectory from point K^* to point F - *Figure 4*.

2.1 The term of 'layer weights'

For the sake of easier consideration of the effect of the parameters of meteorological elements -(***me***) on the flight of the projectile -[7,8], let the height of the vertex of the trajectory Y_s be divided into two layers of equal thickness - '**1**' and '**2**', *Figure 5*.

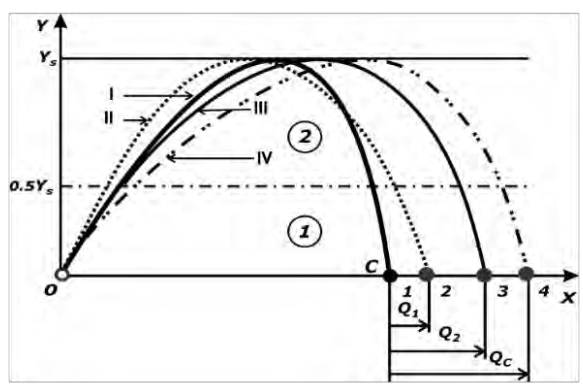


Figure 5. Meteorological elements -(**me**) effect

If ***me***, let's assume that it is the longitudinal wind W_X =0, -nominal, then the trajectory is of the form '**I**' and coincides with the calculated one, so the projectile will be at point '**1**' (**C**) in the horizon. If ***me*** still remain nominal, and W_X first change by 10 units in layer '**1**' - path '**II**', and then in layer '**2**' - path '**III**', the projectile will be at point '**2**' and '**3**'. It can be seen from *Figure 1*.

that the distances Q_1 and Q_2 from point '**1**' (**C**) are not the same because ***me*** are not the same in layers '**1**' and '**2**' either, and this is because their influence in the layers is different. In the event that W_X of 10 units acts simultaneously in layers '**1**' and '**2**' - trajectory '**IV**', the projectile will be at point '**4**' and will be the distance Q_c from point '**1**'. The change in distance Q_c by its numerical value corresponds to the differential coefficient for W_X or correction of the distance due to W_X given in ***FT***, so it will be

$$Q_c = \Delta X_{\mu_1} + \Delta X_{\mu_2} \quad \text{.....(9)}$$

where the values mean: Q_c -deviation of the projectile trajectory in the horizon from point '**1**' (**C**) due to constant wind (called ballistic) acting in both layers, i.e. on whole trajectory; -distance $\Delta X_{\mu_1}, \Delta X_{\mu_2}$ deviations actual ***me*** as for Q_1, Q_2.

Trajectory vertex in principle can be divided into '*n*' layers and consider the impact W_X (as well virtual temperature τ -*Figure 6.* and side wind W_Z), in the same order as for two layers. Then the deviations of the trajectory from the point '**1**' (**C**) would be $Q_1, Q_2, ..., Q_n$ and at the same time $Q_1 \neq Q_2, ..., Q_{n-1} \neq Q_n$.

Sum of the individual deviations gives in the case of a linear influence (in reality it is approximately linear):

$$\sum_{i=1}^{n} Q_i = Q_c \quad \text{.....(10)}$$

Dividing (10) by Q_c yields

$$\sum_{i=1}^{n} \frac{Q_i}{Q_c} = 1 \quad \text{.....(11)}$$

The rations Q_i / Q_c are called "layer weights" and are signed with q_i. The sum of the "weight layers" is equal to one:

$$\sum_{i=1}^{n} q_i = 1 \quad \text{.....(12)}$$

The same conclusion is valid to the "layer weights" of temperature τ and side wind W_Z.

2.2 Weight function

Let the sum of the 'weights layers' q_i from the horizon Y_i to some altitude be denoted with r_i.

Then it is:

For the first layer $r_1 = q_1$;

For the second layer

$$r_2 = q_1 + q_2 = r_1 + q_2 \quad \text{.....(13)}$$

For the it'h and n'th layers $r_i = r_{i-1} + q_i$; $r = r_n = r_{n-1} + q_n$, and the sum of all r_i is:

$$r_n = \sum_{i=1}^{n} q_i = 1 \quad \text{.....(14)}$$

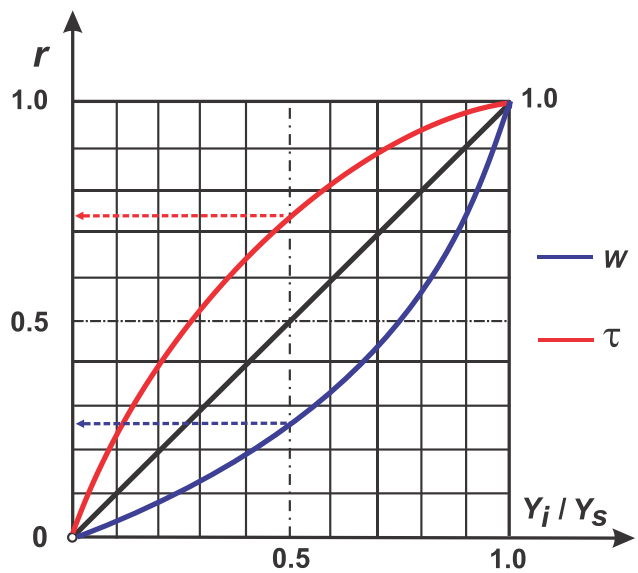


Figure 6. Graph of the 'weight function'

If the obtained r_i values are applied to the graph as a function of the height ratio Y_i / Y_s, where Y_s is trajectory vertex, and Y_i the altitude is from the horizon to the upper limit of the layer, a graph of the weight function is obtained - *Figure 6.*

When the graph of the 'weight function' is available, it can be used to find the weight of a layer such as

$$q_i = r_{Yi} - r_{Yi-1} \quad \text{.....(15)}$$

Where: - q_i is the 'weight' of the layer between Y_i and Y_{i-1}, and r_{Yi}, r_{Yi-1} are numerical values of the 'weight

functions' for the ratios of the altitudes $(0-Y_i)/Y_s$ and $(0-Y_{i-1})/Y_s$.

From *Figure 4*, it can be seen that for $Y_i/Y_s = 0.5$-up to half the trajectory vertex Y_s, the influence of constant (ballistic) wind is about 0.27% of the total effect on the entire trajectory, while for temperature it is about 0.73%. These numerous amounts are obtained for a specific projectile and are not the same for each other, but the qualitative result is similar. It is of key importance when it comes to unguided projectiles: - the influence of temperature (virtual) disturbances is more significant in the ground layers $(0-Y_i)$ than the influence of $(Y_i - Y_s)$, and the opposite is true for the influence of the longitudinal wind W_X. Figure 6. serve as an orientation which ***me*** to choose in which period of time - season, although the effect of the others is cumulative.

2.3 Predicting procedure

The procedure for predicting location where the projectile will fall is as follows:

1. For $(W_X)_{10}$-10 units, the elements of the trajectory to the altitude $Y_{K*=K}$ are calculated, the length $K*-K=\delta_{10}$ is obtained;

2. Since nominal X_K is known and too measured distance X_{K*}, difference is signed as δ_X;

3. The required wind speed W_X is obtained (as well virtual temperature τ and side wind W_Z) from the relation:

$$\frac{\delta_X}{\delta_{10}} = \frac{W_X}{(W_X)_{10}}; W_X = \frac{\delta_X}{\delta_{10}}(W_X)_{10} \quad \text{.....(16)}$$

With determined actual W_X according to (16), by chosen 6-DOF can be calculated actual trajectory parameters. Six degrees of freedom (6-DOF) -[2] is used for long-range trajectories of spin-stabilized projectiles via atmospheric flight to final impact point. Trajectory calculation ought to be performed by 6-DOF ***de*** from $0-F$ -*Figure 4,5*. The 6-DOF ***de*** model was carefully selected for these purposes - [9,10], modified existing -[2] according to -[11], and replaced software of the existing ***BM***.

3. VERIFICATION VALIDITY OF PREDICTION METHOD

In principle, model verification is very simple. The measured numerical values at impact point are simply compared with predicted. If the differences are within acceptable limits, the method is adequate.

Picture 1. Fire correction sonde

The reason that this phase was not completed is the finished design of the hardware placed in the projectile fuze - *Picture 1*.

This device is identical in shape and dimensions to the fuze and is called a *fire correction sonde*. The developing of test equipment for work sonde simulation is to come, and then follow testing by shooting at the proving ground.

It is planned that the description of the hardware and other devices accompanying it will be given in a separate article, which is currently being developed. By carefully observing the sonde, two colors can be observed. The darker one refers to the metal part, and the gray layer on the metal for antenna purposes.

4. CONCLUSIONS

Long time before and nowadays in Artillery fire (gun, howitzer) with spin stabilized projectiles is in use the so-called Corrective shooting procedure. It is given new method based on firing first projectile, without classical

observing its location point of impact in relation to the target. Further follow as usually correcting initial fire elements (aiming) for the second projectile.

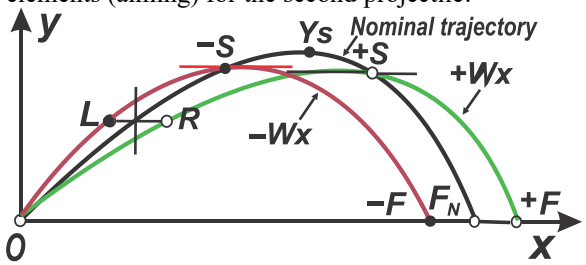


Figure 7. Real trajectories relations

Problem illustrations in question frequently don't exhibit real state of affair nature (*Figure 4*. -literature!). Real state as for relations nominal and disturbed trajectories is exhibited on *Figure 7*. Looking at it can be noticed points $\pm S$ of intersection real trajectories with nominal. Horizontal differences as are "L" or "R" are zero values at it, and developed software for prediction method recognize region where distances "L" - "R" can be used or not and how -[7,8]. Importance of this method is not visible at glance in full scope. One part is explicit, but other, very important as for benefit, is not. Main goal was eliminating long time flight projectile to the target and dangerous long time staying on the fire position. Benefit by the way is that atmospheric parameters used for predicting impact point can be used as knowledge of atmospheric state as for wind, or temperature. The paper presents method and the procedure for predicting the location of the projectile's impact point related to the distant target from a gun. It is based on the measurement of projectile position in space at the initial part of the trajectory, before it reaches the trajectory vertex. Key factor on which method is based is horizontal distance K^*-K - *Figure 4 ("L"-"R" Figure -7!).* By inspection, it is found out ballistic wind value W_X that disturb nominal trajectory at the altitude $Y = Y_K$. After that phase, by ***dj*** renewed ***BM*** are calculated parameters of disturbed trajectory different from nominal one.

References

[1] Petrović, D., ‚Algorithms for predicting the solution of the inverse external ballistic problem', Elaborat VTI-02-01-0866, Belgrade, 2004.

[2] Petrović, D., ‚Prediction solution of inverse External Ballistics problem for Fire Control System Ballistic Module', 1. Naucni Skup Odbrambene Tehnologije u funkciji mira, Beograd, 6-7.12.2005.

[3] Petrović, D., ‚Prediction of artillery correction elements by software for Ballistics Module of fire control systems', 11TH INTERN. SCIEN. CONF. OF DEFANSIVE TECNOLOGIES OTEH 24, Serbia, Tara, 9-11 October 2024.

[4] Apazov R F., Lavrov S. S., Misin V. P., Ballistika upravljae. raket dalnego deistvija, -M, Nauka, 1966.

[5] Islinskiji A. Ju., Inercialnoje upravlenije ballistic. raketami, -M., 1968.

[6] Siharulidze Ju. G., Ballistika letatelnih apparatov, -Moskva, Nauka, 1982.

[7] Petrović D, Analiza meteoroloskih uticaja u Spoljnoj Balistici, Mag. thesis, Beograd, 1982.

[8] Petrović D, Istrazivanje efekata aimetrije u dinamici leta brzo i sporo rotirajucih projektila, Phil. dis., Beograd, 1995.

[9] Abgarjan K. A., Rapoport I. M., Dinamika raket, Masinostroenie, Moskva 1969.

[10] Dmitievskiji A. A., Lisenko L. N., Vnesnjaja Ballistika, Moskva, Masinostroenie, 2005.

[11] Dimitrios N. Gkritzapis, Elias E. Panagiotop., Dionissios P. Margaris and Dimitrios G. Papanikas, A Six Degree of Freedom Trajectory Analysis of Spin-Stabilized Projectiles, International Conference of Comp. Methods in Sciences and Engineering 2007, ICCMSE 2007, Corfu, Greece, 25-30 Sep. 2007.

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION IV

Ammunition and energetic materials - AEM

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

STABILITY EVALUATION OF MULTI-COMPONENT PBX UNDER VARYING CLIMATIC CONDITIONS

MIRJANA KRSTOVIĆ
Military technical institute, Belgrade, mirjana.krstovic@mod.gov.rs

TEODORA STANČIĆ
Military technical institute, Belgrade, teodorastancic@gmail.com

VESNA PETROVIĆ
Technical Overhaul Facility Kragujevac, lab@trzk.co.rs

KATARINA NESTOROVIĆ
Military technical institute, Belgrade, katarina.nestorovic@gmail.com

JELENA ŠULTANS
Technical Overhaul Facility Kragujevac, lab@trzk.co.rs

DANICA BAJIĆ
Military technical institute, Belgrade, danica.bajic@mod.gov.rs

Abstract: *Polymer-bonded explosives (PBX) are widely recognized as reduced-sensitivity energetic materials with improved safety characteristics and satisfactory stability during prolonged adequate storage. However, despite their favorable stability characteristics, the influence of different climatic conditions on their long-term behavior has not been sufficiently investigated, especially when it comes to complex multi-component formulations, like thermobaric (TBX). Systematic studies investigating the combined effects of temperature, humidity, and/or corrosive marine atmospheres on stability of these formulations remain limited. In this paper, multi-component PBX based on hydroxyl-terminated polybutadiene-based binder, containing octogen as the energetic filler, ammonium perchlorate as oxidizer and aluminum and magnesium metal powders were examined. The samples were subjected to controlled environmental conditioning designed to simulate three distinct climatic regimes: continental climate conditions, elevated temperature and humidity conditions, and Mediterranean climate conditions involving exposure to salt-mist to reproduce a marine corrosive atmosphere. The treated samples were evaluated using the vacuum stability test STANAG 4556 procedure and by heat flow calorimetry. The combined application of standardized stability testing and sensitive thermal analysis enabled a comprehensive assessment of climate-induced changes in PBX behavior. The results provide insight into the stability-related responses of PBX under diverse environmental stresses and contribute to a better understanding of their performance and reliability in different climatic deployment scenarios.*

Keywords: *PBX, Vacuum Stability Test, Heat flow calorimetry, environmental conditioning*

1. INTRODUCTION

Cast-cured polymer bonded explosives (PBX) have attracted significant interest due to their reduced sensitivity compared to conventional TNT-based formulations and due to their broad range of detonation characteristics [1]. The main benefits of PBX include significant mechanical strength, increased filling density, reduced sensitivity during handling and storage, good thermal stability and the capability to modulate detonation velocity according to requirements [2]. The potential to integrate metal powders into compositions to produce a thermobaric effect, specifically generating a fireball by the combustion of metal powders post-detonation [3-5], highlights the increase application of these compositions in the weapons industry. Such compositions provide significantly enhanced efficiency in both enclosed spaces, such as tunnels and caves, and in open, cleared areas.

Although many PBX formulations have been developed and generally show low sensitivity to external influences, the effect of environmental factors on their long-term behavior is still not widely studied. There are several reasons for that: the explosives are stored under controlled conditions and sealed in hermetically closed warheads, which limits their exposure to the environment. The extensive variety of munitions that can incorporate these formulations, coupled with their diminished sensitivity, allows for the development of new munitions that, under storage conditions, would not necessitate complete hermetization (for instance, an opening in the liner for the insertion of explosives and placing fuze) until the time of usage. Due to these potential designs, it is essential to

Corresponding author: M. Krstović;
E-mail: mirjana.krstovic@mod.gov.rs
DOI: 10.5937/OTEH26041K

gather information concerning the behavior of these compositions under various climatic conditions, corresponding to different types of temporary storage and exploitation.

This paper examines three samples of multi-component thermobaric PBX explosives that were manufactured five years ago and stored in continental climate conditions during that time. To replicate various climatic events, the samples were subjected to a regime of elevated temperature and air humidity in the climate chamber, along with a regime of high temperature and salt mist to imitate extreme climatic conditions. The explosives were further evaluated using vacuum stability tests and heat flow calorimetry to determine their chemical stability and to acquire a broader understanding of their behavior under extreme environmental conditions.

2. EXPERIMENTAL

Samples of PBX were selected, that represent promising thermobaric compositions with small differences in mass ratio of the ingredients. Formulations with aluminum are very widespread because its characteristics (high heat of combustion, cost and availability [3]), while the Al-Mg combination is used because Mg lowers the ignition temperature and improves reactivity [6].

2.1. Materials

TBX samples were prepared based on:

- Cyclotetramethylenetetranitramine (HMX) from "Prva Iskra Namenska", Barič, Serbia
- ammonium perchlorate (AP), from "Poliex", Berane, Montenegro
- aluminium powder (Al), from "SRCT", China
- magnesium powder (Mg), from "Jengo Spa", China
- binder: hydroxyl-terminated polybutadiene, HTPB ("Polyvest HT Evonik") cured with isophorone diisocyanate, IPDI ("Sigma Aldrich"), and plasticized with dioctyl adipate, DOA.

The appearance of the individual components that are part of TBX formulation can be seen in Figure 1.

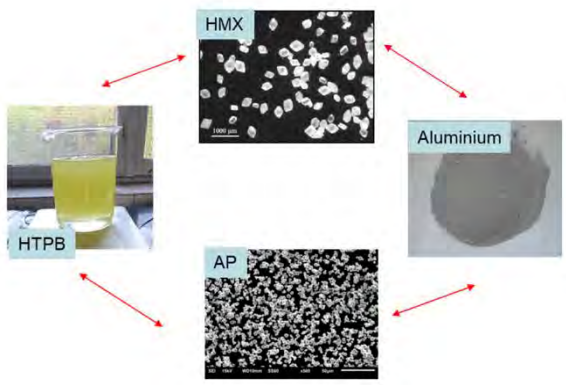


Figure 1. Samle ingridients: HTPB, HMX, Al and AP

2.2. Sample preparation

All the samples have been produced in a vertical homogenizer at 60°C under vacuum according to the process specified by the Military technical institute. Three sample types have been manufactured from three distinct batches, two have the same formulation, while the third has two types of metal powders. All the samples were cast and cured in heating chamber at 70°C during 5 days. The analyzed samples compositions were as follows:

TBX 1: HMX/Al/AP/HTPB;
TBX 2: HMX/Al/AP/HTPB;
TBX 3: HMX/Al/Mg/AP/HTPB.

After the production process, samples from each batch were poured into a vessel with an opening, cured and stored, to provide complete exposure to environmental conditions in storage on one side.

The samples are divided (cut) into three parts: one represents a sample exposed to natural aging, one to be additionally exposed to accelerated aging in continental conditions, and last one to accelerated aging in Mediterranean conditions. The first part, without any additional treatment, was subjected to stability tests and is a reference sample naturally aged in a warehouse for five years in continental climate conditions. The second part was exposed to the process of accelerated aging in the conditions of a continental climate, which was simulated in a climate chamber with an elevated temperature of 60°C and air humidity of 65%, which can be seen in Figure 2.

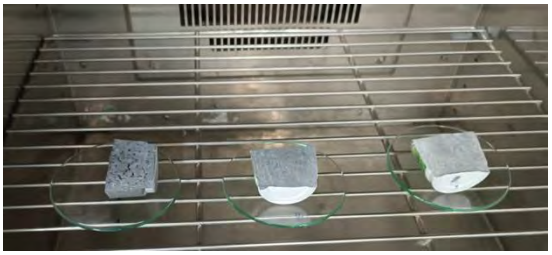


Figure 2. TBX in climate chamber (from left: TBX 3, TBX 2, TBE 1)

The third part of the samples was exposed to extreme conditions - a temperature of 30°C and air humidity of 90% in conditions of marine atmosphere imitated by a salt-mist camber. The sample that was in the salt-mist chamber for 48 hours, after the experiment, due to extremely high humidity, was covered with salt water that condensed on the lid of the chamber during the operation of the chamber and dripped on the samples.

After treatment, all samples were mechanically powdered to particles of approximately 2×2 mm (Figure 3) in order to prepare them for further analysis.

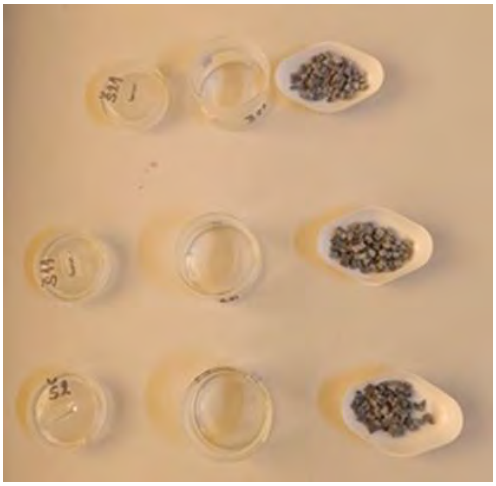


Figure 3. Chopped samples of naturally aged TBX

The moisture content of each sample had been previously determined using the drying method at 10

0 °C for 2 h, in order to evaluate the influence of moisture on the treated samples.

2.3. Methods

Chemical and thermal stability of the investigated materials was evaluated using the STABIL system (modernized vacuum stability tester) apparatus manufactured by OZM Research, Czech Republic. Method was performed by STANAG 4556 [7,8] procedures, by using 2.5 g of sample. The chopped samples are placed in a test tube (Fig. 4a) and heated at 100 °C for a duration of 40 h.

Figure 4. a) Prepared sample of TBX in test tube for VST and b) prepared sample of TBX in vial for HFC

Heat flow calorimetry was method of choice because it can also yield information about the stability of energetic materials measuring the decomposition rate. Experiment was carried out by at 85°C and using Heat Flow Calorimeter, TAM IV TA Instruments with 24 channel for 7 days. Approximately 3g of the chopped sample was placed in the vial for microcalorimetry (Fig. 4b).

3. RESULTS AND DISCUSSION

All samples undergo the same type of tests, with the same conditions to obtain the comparable results in stability.

3.1. Vacuum stability results

The moisture level of the surface and middle layers was initially assessed, and the results of moisture content are given in Table 1. Vacuum stability test at first was carried out for both layers for naturally aged samples to detect the slight differences in moisture content and the gas evolved in the vacuum stability test tube for the two layers are shown in Table 2.

Table 1. Moisture content for naturally aged samples

Sample	naturally aged		clime chamber		salt-mist chamber	
	surface layer (mass%)	middle layer (mass%)	surface layer (mass%)	middle layer (mass%)	surface layer (mass%)	middle layer (mass%)
TBX 1	0.02	0.04	0.04	0.05	0.04	0.04
TBX 2	0.00	0.04	0.04	0.04	0.04	0.04
TBX 3	0.28	0.41	0.14	0.08	0.06	0.12

Table 2. Vacuum stability test result for naturally aged samples by layers

Sample name		Gas evolved [cm3/g]		
		test tube 1	test tube 2	average
surface layer	TBX 1	0.115	0.118	0.117
	TBX 2	0.116	0.117	0.117
	TBX 3	0.398	0.404	0.401
middle layer	TBX 1	0.100	0.101	0.101
	TBX 2	0.106	0.108	0.107
	TBX 3	0.364	0.364	0.364

Table 2. demonstrates that all samples exhibit almost full resistance to moisture. The elevated value primarily reflects the formulation including a bi-component mixture of fuel, i.e. aluminum and magnesium. According to the fact that the samples were not hermetically sealed during storage, and had an opening on one side, it was necessary to consider the influence of moisture. Considering that the trend remains the same across all compositions and that similar values are observed in different layers and within the same samples for both moisture and released gases (Table 3) during the stability test, the remaining samples will be evaluated only in one middle layer and compared to that layer of naturally aged samples.

Samples exposed to the climate chamber under conditions of elevated humidity and temperature had their moisture content assessed, whereas undried samples were submitted to the vacuum stability test. The obtained results are given in Table 3.

Table 3. Vacuum stability test result for climate chamber samples

Sample name	Gas evolved [cm^3/g]		
	test tube 1	test tube 2	average
TBX 1	0.156	0.147	0.151
TBX 2	0.144	0.134	0.139
TBX 3	0.348	0.346	0.347

Extreme conditions, even in a relatively short time interval, lead to changes in the amount of gas released. The compositions containing only Al show an increase of as much as 30 to 50%, while the composition containing the combination of Al and Mg after exposure to the climate chamber gives a lower value by about 5%. However, it is important to note that even those, at first glance extremely high values, are not even close to critical values, which range from 1.2 to 2 cm^3/g according to test criteria [10].

Table 4 summarizes the results acquired from compositions subjected to salt mist. Despite being completely submerged in salt water, the sample's minimal moisture content indicates that the cross-linking of the HTPB with IPDI produced a sufficiently resilient rubbery binder structure that exhibits remarkable resistance to moisture and water.

Table 4. Vacuum stability test result for salt-mist chamber samples

Sample name	Gas evolved [cm^3/g]		
	test tube 1	test tube 2	average
TBX 1	0.195	0.198	0.172
TBX 2	0.198	0.199	0.199
TBX 3	0.331	0.321	0.326

Considering the composition of TBX 1, in each of the mentioned atmospheres, there is a clearly noticeable trend of change in the amount of evolved gases in rapidly ging samples in clime and salt-mist chambers. In Figure 4 we may observe the differences between atmospheres.

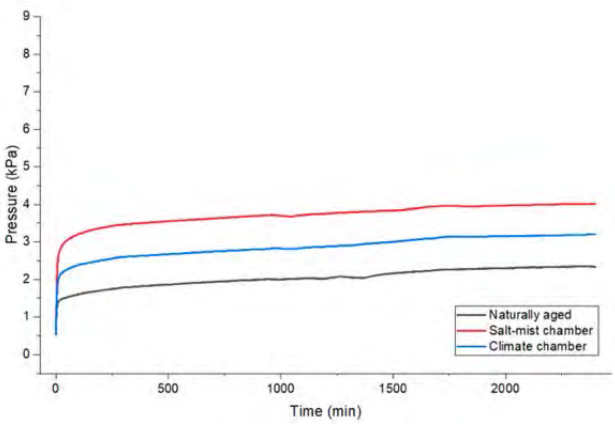


Figure 5. Sample of TBX 1 in different atmospheres

Figure 5 presents shows that all samples exhibit a rapid initial pressure increase during the early stage of the test, followed by a gradual stabilization phase. The sample exposed to the salt-mist chamber shows the highest-pressure values throughout the entire testing period, reaching approximately 4.0 kPa while climate chamber sample stabilizing around 3.1–3.2 kPa. Naturally aged samples exhibit the lowest pressure development over time, reaching approximately 2.3 kPa.

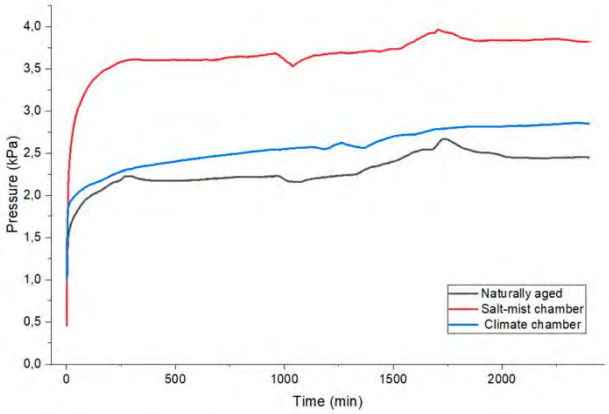


Figure 6. Sample of TBX 2 in different atmosphere

The differences between the samples are obtained almost immediately, the curves reach maximum values in first minutes of the test. The curve shapes indicate stabile behavior during test period, without any sharp picks which could reveal instability. Although this finding proving chemical stability providing extremely low values of evolved gas compared to German test criteria, it emphasizes significantly environmental influence on samples [11,12]

A very similar trend was recorded in sample TBX 2, which is identical in composition to TBX 1. The values of the naturally aged sample and the sample from the air conditioning chamber are somewhat closer, while there is a big noticeable difference in the sample exposed to the simulation of the Mediterranean climate.

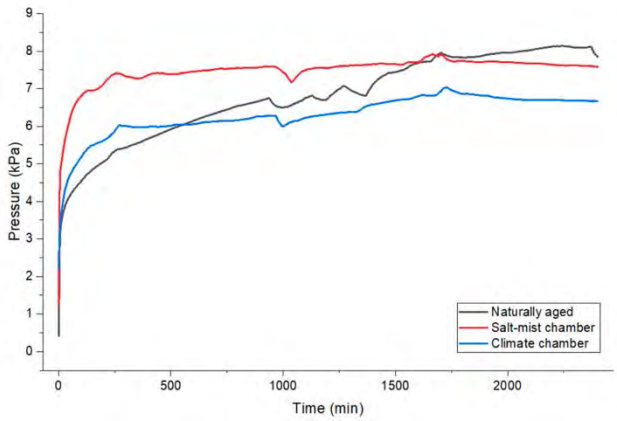


Figure 7. Sample of TBX 3 in different atmosphere

The sample TBX 3, shown in Figure 7 clearly indicate that the behavior of composites that contain a combination of two metal powders, Al and Mg, does not follow the trend observed in samples TBX 1 and TBX 2. Samples of TBX 3 from all three atmospheres show a higher sensitivity to the stability test, which in some

segments is even three times higher compared to equally treated samples that do not contain magnesium. However, a greater resistance to the influence of the atmosphere is also clearly noticeable, that is, all three samples show very similar behavior and reach close values.

Also, it can be detected that the pressure of naturally aged sample changes linearly with time, while the samples exposed to the climate chamber and salt-mist chamber after the initial pressure change, keep their values stable. The values obtained for the magnesium-containing compositions indicate a certain variation in stability over time. Although all measured parameters remain well below the prescribed limit values, it is evident that their thermal stability behavior differs compared to TBX 1 and TBX 2. This deviation does not necessarily imply instability; however, it suggests that the presence of magnesium influences the thermal response of the system. Therefore, these compositions require continued monitoring and additional investigation to fully understand their long-term stability and performance characteristics.

3.2. Heat flow capacity

Considering that the examined PBXs have different components, the behavior of each individual sample exposed to different treatment was monitored using the HFC method. Table 5 provides an overview of the measured maximum heat flow for all samples.

Table 5. Maximum heat flow for TBX 1-3

Samp.	treatment	Max heat flow (µW/g)		
		sample 1	sample 2	Average
TBX1	naturally aged	4.26	4.27	4.26
	climate chamber	3.66	4.02	3.84
	salt-mist chamber	4.09	4.09	4.09
TBX2	naturally aged	3.42	3.52	3.47
	climate chamber	3.67	3.65	3.66
	salt-mist chamber	5.07	4.35	4.71
TBX3	naturally aged	218.3	214.9	216.6
	climate chamber	183.3	194.5	188.9
	salt-mist chamber	257.0	280.8	268.9

The table reveals that samples TBX 1 and TBX 2 have comparable results that are significantly below the thresholds established by STANAG 4582 (for 85°C for 5,98 days heat flow limit is 201µW/g). It is important to note that the document addresses gunpowder and rocket fuels [8,9], which are susceptible to thermal breakdown and self-ignition under critical conditions, but in absence of adequate standard method for PBX, this criteria was chosen for the examined explosives as well. However, the thermal activity shows significantly different results when the TBX 3 sample is taken into consideration. Its values slightly exceed the allowed limit, which indicates that should conduct additional analyzes and introduce other methods that would give additional insight into the thermal stability of the mentioned compositions.

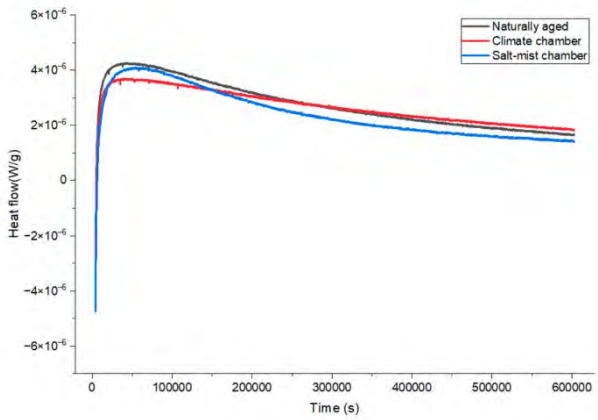


Figure 8. HFC diagram for sample TBX 1

As it can be seen in figure 8, all treated samples of PBX 1 exhibited an initial exothermic peak followed by a gradual decline in heat flow over time. The maximum heat flow values followed the order: naturally aged, salt-mist chamber and climate chamber. The naturally aged sample showed the highest peak intensity, indicating the greatest initial reactivity. The salt-mist exposed sample displayed a slightly lower maximum, while the climate chamber sample exhibited the lowest heat flow. The initial exothermic peak could be potentially influenced by residual moisture since the samples were not pre-dried. Following the maximum, all samples exhibited a similar decay trend, without pronounced and low endothermic or exothermic peaks with a tendency to decrease thermal activity over time.

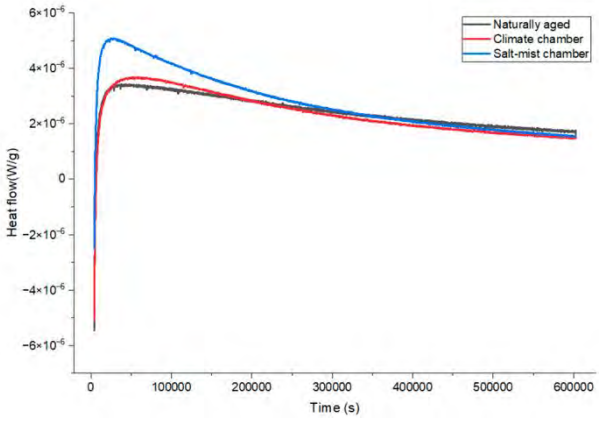


Figure 9. HFC diagram for sample TBX 2

At Figure 9, for TBX 2 the heat flow curves also showed a pronounced initial exothermic peak followed by a gradual decay for all three conditioning treatments. In contrast to the PBX 1 sample, the maximum heat flow has a different trend of thermal response with highest values for salt-mist treated sample, the climate chamber sample shows intermediate reactivity, while the naturally aged specimen presents the lowest maximum heat flow. The sharper and higher peak observed for the salt-mist sample suggests enhanced initial reactivity which can be caused both by oxidation and moisture-related processes.
After the maximum, all curves exhibit a similar long-term decay trend, indicating progressively and limited reactions.

Compared to Figure 8, a different trend can be noticed and clear reversal in maximum heat flow. The unexpected variations in thermal response among samples of identical composition can be attributed to their derivation from distinct batches, wherein the raw materials can possess difference pretreatment processes, particularly at elevated temperatures, such as those employed during the HFC test. Overall, the comparison indicates that environmental conditioning significantly alters the initial kinetics, while long-term behavior remains governed by similar diffusion-controlled mechanisms.

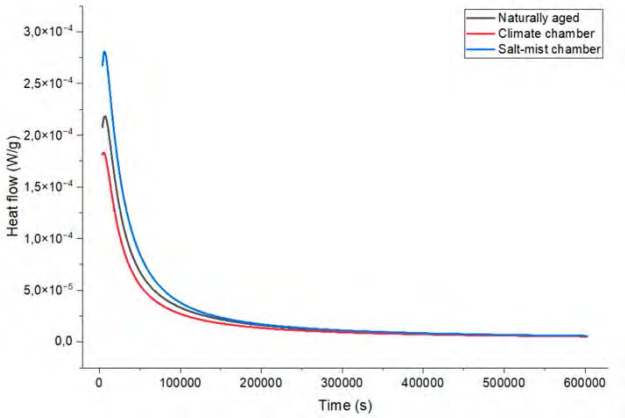


Figure 10. HFC diagram for sample TBX 3 in different atmospheres

However, regarding the PBX 3 sample, it is an entirely different sample. Specifically, the maximum heat flow is up to 50 times higher than in the formulation that comprises only Al as the fuel component.

All tree samples show very intense initial exothermic response followed by a rapid decay toward near-zero heat flow. The peak occurs very early and is sharp, indicating a highly reactive initial process. After 28 hours all curves converge and approach very low heat flow values, suggesting depletion of reactive processes. The salt-mist sample exhibits the highest peak, while the climate chamber sample consistently shows the lowest reactivity. In all three cases the rapid decay suggests that the process is primarily surface-controlled rather than diffusion-controlled over long times, according to the fact that the tested samples were not previously dried, so the initial reactions may be the result of a small residual amount of moisture in the samples

Analyzing all three samples, it is clearly observed that there are differences in the thermal response of samples exposed to certain conditions of high temperature, humidity or salt mist, which indicate that environmental conditioning significantly affects thermal stability of the examined PBX.

Salt-mist exposure tends to enhance initial reactivity, which can be seen particularly evident in the second and third diagram, while climate chamber exposure generally reduces maximum heat flow.

Despite differences in peak magnitude, all datasets eventually converge to low heat flow values, indicating exhaustion of reactive ingredients and transition to kinetically limited process. This behavior indicates pronounced short-term temperature sensitivity. It is important to note that PBX compounds are not exposed to excessively high temperatures during production, as mixing and crosslinking processes do not exceed 70 °C. Furthermore, the operational temperature range of ammunition is relatively moderate, typically between −30 °C and +50 °C. The HFC diagrams, together with the maximum heat flow values, demonstrate that all detectable processes, whether associated with the release of residual moisture, reactions with oxygen present in the vial, or relaxation of the polymer matrix, occur within a limited time interval no longer than 24 hours. After this initial period, the heat flow rapidly decreases, indicating that the system reaches a relatively stable state with no sustained long-term exothermic activity.

4. CONCLUSION

This study evaluated the environmental and thermal stability of three PBX formulations subjected to natural aging, elevated temperature and humidity, and salt-mist exposure to simulate Mediterranean climate. Vacuum stability testing demonstrated that all compositions released gas amounts significantly below the critical limits defined by 1.2 cm^3/g, confirming their overall chemical stability. Nevertheless, environmental conditioning influenced early gas evolution kinetics. Salt-mist exposure produced the highest-pressure development, while naturally aged samples consistently exhibited the lowest values. It is also noticed that compositions containing only aluminum (TBX 1 and TBX 2) showed comparable and stable behavior for all conditioning treatments. In contrast, TBX 3, which incorporates a combination of aluminum and magnesium powders, revealed increased sensitivity in both vacuum stability and heat flow measurements. Although it exhibited acceptable safety in vacuum stability testing, TBX 3 exhibited substantially higher thermal activity. Heat flow capacity results confirmed that TBX 1 and TBX 2 remained far below the STANAG 4582 heat flow limit, whereas TBX 3 exceeded the prescribed threshold, indicating the need for further thermal characterization.
Environmental exposure primarily affected early-stage reaction kinetics, particularly enhancing reactivity after salt-mist conditioning, while long-term behavior converged toward low heat flow values for all formulations. The results confirm that the cross-linked polymer matrix provides effective resistance to moisture and atmospheric influence. However, the presence of magnesium significantly alters thermal response, emphasizing the importance of continued monitoring and additional long-term stability investigations for Mg-containing formulations. Also, the findings encourage further research in establishing criteria and finding appropriate methodology of testing the long-term stability of multi-component PBX.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325, and University of Defense, Military Academy, Proj. No. VA-TT/1/22-24.

References

[1] SIMIC D: *Liveni termobarični PBX eksplozivi*, Vojnotehnički institut Beograd, Naučnotehničke informacije, Vol53, No. 2, ISSN 1820-3418, 2016.

[2] ORBOVIC N. B.: *Uvod u energetske materijale, hemijski aspekti razlaganja energetskih materija*, KIZ CENTAR, Beograd, 2020.

[3] TÜRKER L.: *Thermobaric and enhanced blast explosives (TBX and EBX),* Defense Technology 12 (2016), pp. 423–445

[4] TRZCINSKI W. A., MAIZ L.: *Thermobaric and enhanced blast explosives – properties and testing methods,* Propellants, Explosives, Pyrotechnics, 2015, Vol.40, No.5, pp.632-644.

[5] KRSTOVIC M., BAJIC D., TIMOTIJEVIC M., NESIC J., FIDANOVSKI B.: *The influence of nano-scale aluminum on the energetic materials – theoretical and experimental observations*, Scientific Technical Review, Vol. 74, No. 1, 2024. pp. 36-42

[6] S. CUDZIŁO, W. A. TRZCIŃSKI, J. PASZULA, M. SZALA, Z. CHYŁEK: *Performance of Magnesium, Mg-Al Alloy and Silicon in Thermobaric Explosives – A Comparison to Aluminum*, Prop., Explos., Pyrotech. 2020, 45, 1691.

[7] STANAG 4556, *Explosives: Vacuum Stability Test, (Edition 1)* 22 November 1999; MAS/1436-PPS/4556; North Atlantic Treaty Organization.

[8] KRSTOVIĆ M., STANČIĆ T., NESTOROVIĆ K., BAJIĆ Z., BOGDANOV J., BAJIĆ B.: *Vacuum stability evaluation of pyrotechnic mixture - double-base propellant compatibility*, New Trends in Research of Energetic Materials, University of Pardubice, Pardubice, 22-24 April 2026, pp. 426-435.

[9] Ek S., Sjöblom M., Brantlind M., and Johansson J.: *Synthesis and characterization of the energetic plasticizer Bu-AENA,* New Trends in Research of Energetic Materials, University of Pardubice, Pardubice, 2-4 April 2025, pp. 109-117.

[10] Bohn, A. M: *NC-based energetic materials – stability, decomposition, and ageing, Presentation on the Meeting Nitrocellulose – Supply*, Ageing and Characterization, April 24 to 25, 2007, AWE, Aldermaston, England

[11] SANZHEN W., MINGKUN F., XINGLIANG W., GUANGFEI G., JUNHONG W., SEN X.: *Effects of temperature, particle size, and air humidity on sensibility of typical high-energetic explosives*, Journal of Measurement Science and Instrumentation, 2024, 15（3）：pp. 408-416.

[12] HAOYING L., LI H., YANCHUAN J. AIFENG J., DONGMING S.: *The influence of temperature and humidity on the decomposition of HMX-HTPB polymer-bonded explosives*, Thermochimica Acta, Volume 749, 2025,

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EXPERIMENTAL AND NUMERICAL STUDY OF PRECURSOR SHAPED CHARGE PERFORMANCE

PREDRAG ELEK
University of Belgrade – Faculty of Mechanical Engineering, Weapon Systems Department, Belgrade, pelek@mas.bg.ac.rs

BILJANA JOVANOVIĆ
University of Belgrade – Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Department of Physics, Belgrade, jbiljana@vin.bg.ac.rs

DRAGANA JELIĆ
HI Poliex AD, Research and Development Center, Berane, Montenegro, jelicdragana77@gmail.com

DANIJELA OBRADOVIĆ
HI Poliex AD, Research and Development Center, Berane, Montenegro, dobradovic371@gmail.com

DANICA JELIĆ
HI Poliex AD, Research and Development Center, Berane, Montenegro, djelicpoliexgroup@gmail.com

DUŠAN BLAGOJEVIĆ
Forging Center LLC - Valjevo, Valjevo, dusan.blagojevic@kovacki-centar.com

Abstract: *The shaped-charge effect remains one of the most effective mechanisms for defeating armored targets. Tandem shaped-charge warheads provide enhanced effectiveness against explosive reactive armor systems, where the precursor charge is used to initiate the reactive armor prior to the arrival of the main charge. The present study investigates the performance of a small-diameter precursor shaped-charge warhead with a charge diameter of 50 mm. A 60° conical copper liner with a thickness of 1 mm is accelerated by the detonation products of an explosive charge and subsequently transformed into a high-velocity jet. The experimental program included three explosive materials—TNT, RDX, and HMX—as well as variations in standoff distance. A high-strength alloy steel cylinder was used as the target. In parallel, a numerical model of jet formation and target penetration was developed using the commercial finite element software Abaqus/Explicit and a purely Eulerian formulation. Comparison of the experimental observations and numerical predictions demonstrated good agreement in terms of penetration depth and cavity geometry, while also providing insight into the underlying penetration mechanisms and the influence of key design parameters. The validated numerical model provides a reliable basis for further analysis and optimization of shaped-charge warhead performance.*

Keywords*: precursor shaped charge, shaped charge jet, penetration depth, experimental investigation, numerical simulation*

1. INTRODUCTION

More than a century after its discovery, the shaped-charge effect remains one of the most effective mechanisms for armor penetration and continues to serve as the basis for numerous modern warhead designs. Its operating principle relies on focusing explosive energy and converting it into a high-velocity metal jet through the collapse of a metallic liner [1–5]. Defense applications have driven the continuous improvement of shaped-charge performance through extensive experimental, analytical, and numerical investigations.

The shaped-charge mechanism may be divided into the following phases [1,2]: (i) propagation of the detonation wave through the explosive charge, (ii) acceleration of the liner material by the detonation products, (iii) liner collapse and jet formation, (iv) jet evolution and stretching, (v) jet impact on the target and subsequent penetration, and (vi) process termination, either by target perforation or by arrest of penetration in a sufficiently thick target. Table 1 summarizes representative values of the characteristic parameters of shaped-charge warheads and jets, together with the corresponding performance metrics.

In the present study, an experimental investigation of the performance of a small-diameter (50 mm) shaped-charge warhead was conducted. Subsequently, comprehensive numerical simulations of the complete shaped-charge process were performed. The objective of the research was to develop and validate a numerical model capable of

Corresponding author: P. Elek;
E-mail: pelek@mas.bg.ac.rs
DOI: 10.5937/OTEH26042E

accurately reproducing the experimental observations, thereby providing a reliable basis for further optimization of the warhead design and enhancement of its performance.

Table 1. Typical ranges of characteristic parameters and performance metrics for shaped-charge warheads and jets [1,2].

Liner base diameter	*CD* = (40…700) mm
Detonation velocity	*D* = (7.0…10.0) km/s
Liner thickness	(0.02…0.04) *CD*, for copper
Jet diameter	≈ (0.03…0.15) *CD*, variable
Jet length	≈ (6…15) *CD,* before breakup
Cavity diameter	≈ (0.1…0.3) *CD*, variable
Jet tip velocity	≤ 1.4 *D*
Jet tail velocity	≈ (0.2…0.5) *D*
Optimum standoff	≈ (2…7) *CD*
Penetration depth	(4…10) *CD,* in steel target

2. EXPERIMENTAL INVESTIGATION

The subject of the experimental investigation was a small-diameter shaped-charge warhead shown in Fig. 1. The warhead is primarily intended for use as a precursor charge in a tandem-warhead configuration; however, it may also be employed as a standalone warhead carried by an unmanned aerial vehicle (UAV) intended for top-attack engagement of lightly armored targets.

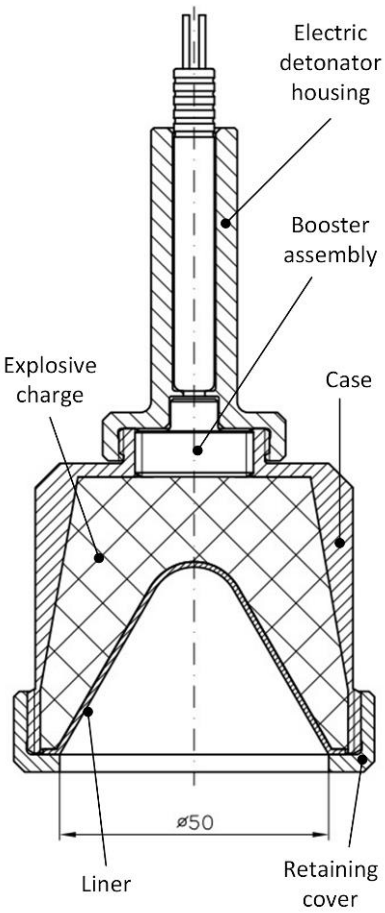


Figure 1. Assembly drawing of the 50 mm shaped-charge warhead configuration employed in static tests, showing its main components.

As a part of a broader experimental campaign, a series of field tests was conducted at the proving ground of Poliex in Berane, Montenegro. In the static tests, the explosive train was initiated using an EDK-8 electric detonator manufactured by Poliex. The detonation impulse was transmitted to a PETN booster charge, ensuring reliable initiation of the main explosive charge. Various explosive formulations were investigated as the main charge material, including phlegmatized RDX, HMX, and TNT. The explosive charges were manufactured by pressing the explosive compositions to the prescribed density.

The conical liner was manufactured at Forging Center LLC Valjevo by cold forging from copper bars supplied by Copper Mill Sevojno, Serbia. The material was DNP-Cu (Class A), a phosphorus-deoxidized copper grade with a copper content of 99.95 wt.% and a phosphorus content below 0.015 wt.%. Following forging and subsequent machining, the liner was placed in a furnace preheated to 300 °C and held at this temperature for 15 minutes. It was then removed from the furnace and allowed to cool in air. The hardness measured immediately after forging, prior to machining and annealing, was 105 HBW 5/250. Due to the small wall thickness of the liner and the limitations of the available hardness testing equipment, its hardness after annealing was not measured.

The warhead case was manufactured from glass-fiber-reinforced polyamide in the baseline configuration, while aluminum and steel were considered as alternative casing materials. The present paper considers only the baseline configuration. The explosive pressing operations, fabrication of individual components, copper liner annealing, and final warhead assembly were all conducted at the Poliex manufacturing facilities. Photographs of the individual components and assembled warheads are presented in Fig. 2.

As a practical substitute for standard rolled homogeneous armor (RHA) steel plates, cylindrical targets made of quenched and tempered Č.4732 steel were employed owing to their availability, cost-effectiveness, and suitability for longitudinal sectioning, which facilitated detailed post-test examination of the penetration cavity. Equivalent designations of this steel in other standards include 42CrMo4 (EN), AISI 4140, and 1.7225 (W.Nr.). The targets had a diameter of 80 mm and lengths of 150 mm and 200 mm.

The standoff distance between the base of the conical liner and the target was controlled using a machined polystyrene spacer designed to ensure accurate alignment and positioning of the warhead relative to the cylindrical target. Adhesive tape was additionally used to secure the assembly during the tests. Photographs of the setup used in the field tests are presented in Fig. 3.

The primary diagnostic parameters considered in the field tests were: (i) penetration depth, (ii) the geometry of the penetration cavity, including the entry and possible exit holes, and (iii) damage to the supporting structure. The experimental results are presented and discussed in Section 4 through comparison with the corresponding numerical predictions.

3. NUMERICAL SIMULATIONS

The numerical investigation of shaped-charge jet penetration was performed using the finite element software Abaqus [6] and its explicit solver. This software has been widely employed for the simulation of transient, high-energy, and high-strain-rate phenomena, including shaped-charge jet formation and penetration into metallic targets [7–11].

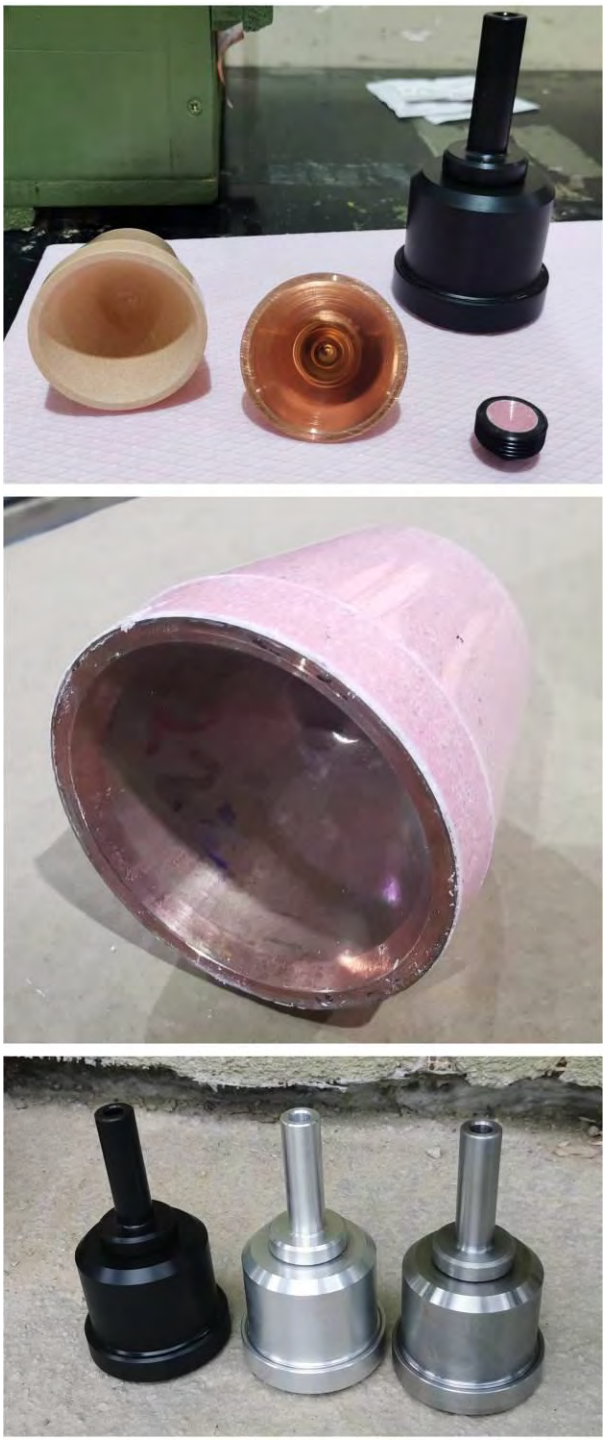


Figure 2. Photographs of the main stages of warhead preparation: (top) individual warhead components, including the RDX explosive charge, liner, booster, and casing; (middle) assembly of the liner and HMX explosive charge; and (bottom) completed warheads with polyamide, aluminum, and steel casings.

Figure 3. Experimental setup: (top) warheads and cylindrical targets configured with various standoff distances; (bottom left) assembled warhead–target configuration before testing and a recovered target showing the penetration cavity; and (bottom right) warhead fitted with an EDK-8 detonator and prepared for testing.

Owing to the severe material deformation associated with the penetration process, a purely Eulerian formulation was employed. A quasi-axisymmetric 15° wedge model was adopted, reproducing the actual geometry of the shaped-charge warhead and target configuration (Fig. 4). Appropriate boundary conditions were applied to enforce symmetry. Following a mesh convergence study, a uniform element size of 0.5 mm was used in the axial direction, while the element size in the radial direction gradually increased from 0.2 mm to 0.8 mm. The total number of Eulerian EC3D8R elements was on the order of 500,000, depending on the model configuration, particularly the target thickness and standoff distance.

Accurate representation of material behavior under highly dynamic loading conditions is of crucial importance for the reliability and predictive capability of the numerical model. The shaped-charge liner and target were modeled using material parameters representative of the copper grade (DNP-Cu, Class A) and Č.4732 steel employed in the experiments, respectively. The dynamic plastic response of both materials is described using the Johnson–Cook constitutive model [12], together with a linear shock-wave equation of state. Material failure is modeled using the Johnson–Cook damage model [13]. The warhead casing was made of glass-fiber-reinforced polyamide (PA-GF), corresponding to the configuration

used in the experimental tests. In the numerical model, the casing material was represented by a simplified isotropic elastic-plastic model with a density of 1350 kg/m³, Young's modulus of 8 GPa, and Poisson's ratio of 0.35. The plastic behavior and ductile damage initiation were described using tabulated parameters estimated based on available literature [14]. Experimental observations and numerical analyses conducted by the authors within broader research program indicated that variations in casing material had a relatively limited influence on the penetration performance of the investigated shaped-charge configurations.

The explosive charge is described using the Jones–Wilkins–Lee (JWL) equation of state [15], coupled with a programmed burn model. The material parameters used in the Abaqus simulations are summarized in Tables 2 and 3.

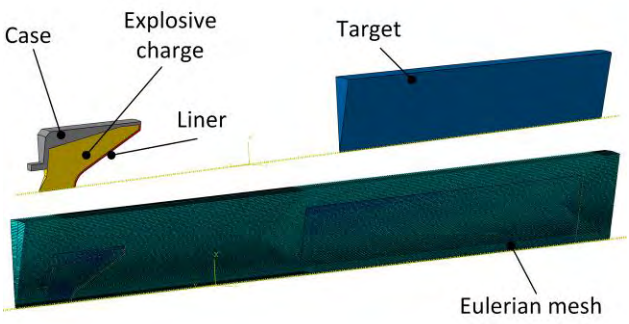


Figure 4. Geometry and Eulerian finite element mesh of the quasi-axisymmetric numerical model.

The total duration of the simulated process, from explosive initiation to the end of jet penetration, is approximately 150–250 µs.

Table 2. Material parameters used in the Abaqus simulations for the copper liner and steel target [12, 13].

	COPPER	Č.4732
Density (kg/m³)	8960	7850
JOHNSON-COOK PLASTICITY MODEL		
A (MPa)	90	792
B (MPa)	292	510
C	0.025	0.014
n	0.31	0.26
m	1.09	1.03
T_{melt} (K)	1356	1800
T_{trans} (K)	300	300
LINEAR U_s – U EQUATION OF STATE		
c_0 (m/s)	6940	4578
s	1.49	1.33
Γ	2.0	1.67
JOHNSON-COOK DAMAGE MODEL		
d_1	0.54	0.05
d_2	4.89	3.44
d_3	3.03	2.12
d_4	0.014	0.002
d_5	1.12	0.61

Table 3. Material properties and Jones–Wilkins–Lee (JWL) equation-of-state parameters for TNT, RDX, and HMX [15–17].

PROPERTY	HMX	RDX	TNT
PHYSICAL AND DETONATION PROPERTIES			
Density (kg/m³)	1760	1600	1550
Det. velocity (m/s)	8667	8112	6800
Det. energy (MJ/kg)	5.55	5.43	4.29
JWL EQUATION OF STATE PARAMETERS			
A (GPa)	674.9	611.3	335.7
B (GPa)	6.130	10.65	2.90
R_1	4.20	4.40	4.15
R_2	1.00	1.20	0.95
ω	0.30	0.32	0.30

4. RESULTS AND DISCUSSION

Figure 5 presents a representative numerical simulation of an HMX-based shaped charge at a standoff distance of 2.6 charge diameters (CD). The sequence of snapshots illustrates the complete process from jet formation and evolution to target penetration and the attainment of the final penetration depth, providing insight into the temporal development of the shaped-charge jet and its interaction with the target.

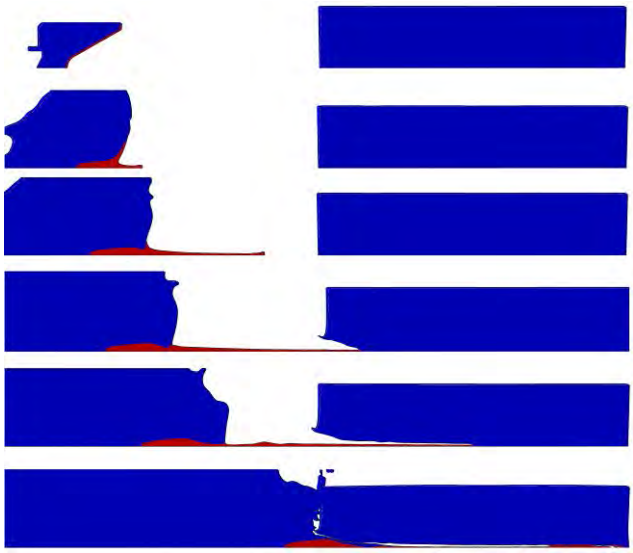


Figure 5. Numerical simulation of shaped-charge jet formation and penetration into a steel target. The snapshots illustrate the temporal evolution of the process at 0, 15, 30, 45, 90, and 240 µs after initiation.

The numerical simulations provide detailed information on the penetration process that cannot be obtained directly from the experiments. For example, the predicted jet tip velocity is approximately 5350 m/s. This relatively low value can be attributed to the large cone angle of the liner, which favors the formation of a heavier but slower jet.

The predictive capability of the numerical model was assessed by comparing the calculated penetration depths for several representative cases with the corresponding experimental measurements, as summarized in Table 4.

Table 4. Comparison of experimentally measured and numerically predicted penetration depths for various explosive charge materials and standoff distances.

Explosive charge	Standoff (CD)	Penetration depth (mm)	
		Simulation	Experiment
HMX	2.6	194	188
RDX	1.2	122	118
TNT	1.2	98	93

Overall, very good agreement between the numerical predictions and the experimental measurements was achieved, with the relative error in penetration depth not exceeding 5%. Interestingly, even the TNT-filled charge provides a significant penetration capability, making it potentially suitable not only for precursor-warhead applications but also as a standalone warhead against lightly armored targets. As expected, the numerical model consistently predicts slightly greater penetration depths, primarily due to the idealized axisymmetric representation of the process, including perfectly symmetric initiation, explosive charge geometry, and liner configuration. These assumptions result in a more coherent and continuous shaped-charge jet than that produced under experimental conditions.

Figure 6 compares the entry holes produced by the HMX-based warhead listed in Table 4, as obtained experimentally and predicted numerically. The comparison demonstrates good agreement between the experimental observations and numerical predictions, confirming the ability of the numerical model to reproduce the principal features of the target response in the vicinity of the entry hole.

After testing, the targets were sectioned longitudinally using wire electrical discharge machining (wire EDM). A comparison between the experimentally obtained penetration cavity profile and its numerical prediction is presented in Fig. 7. The numerically predicted cavity is slightly narrower than that observed experimentally, particularly in the entrance region and over approximately the first third of the penetration depth. This discrepancy can be attributed primarily to the idealized nature of the numerical model, which assumes perfect geometric and loading symmetry and consequently produces a more coherent and better-focused jet. In contrast, the experimental process is inevitably affected by minor geometric and manufacturing imperfections, as well as small asymmetries in initiation and jet formation, which promote jet dispersion and result in a somewhat larger cavity diameter. The agreement improves with increasing penetration depth, indicating that the numerical model captures the overall penetration mechanism and cavity evolution with satisfactory accuracy.

The good agreement between the experimental observations and numerical predictions demonstrates the capability of the numerical model to accurately capture the jet formation and penetration processes, thereby providing a sound basis for further optimization of the shaped-charge warhead design.

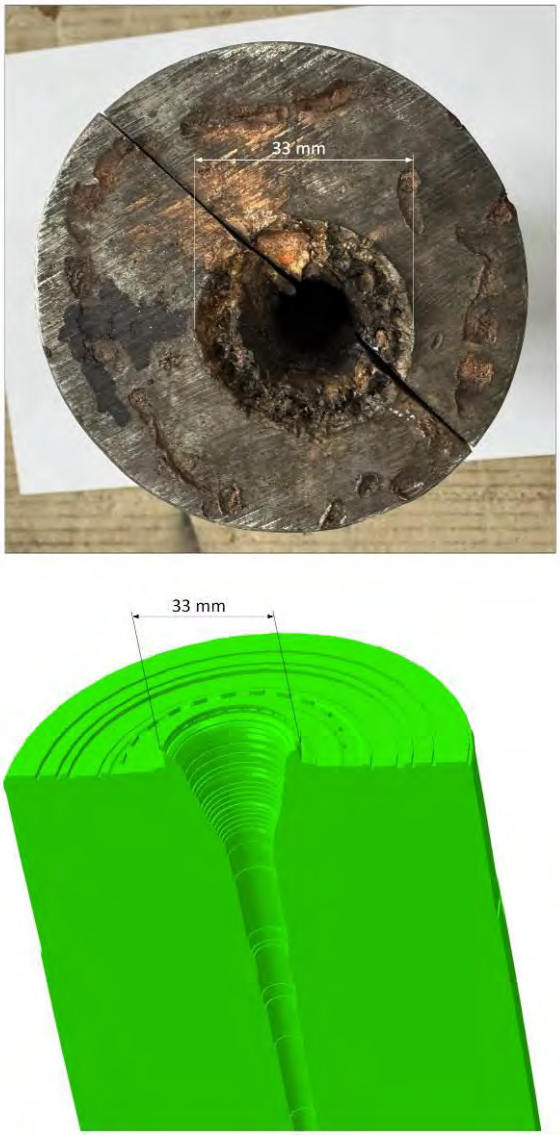


Figure 6. Comparison of experimentally obtained (top) and numerically predicted (bottom) entry holes for an HMX-filled shaped-charge warhead at a standoff distance of 2.6 CD.

5. CONCLUSION

This paper presents an experimental and numerical investigation of a small-diameter shaped-charge warhead with a charge diameter of 50 mm. The field tests examined the effects of explosive charge material and standoff distance using steel cylindrical targets. In parallel, a numerical model of shaped-charge jet formation and target penetration was developed in Abaqus/Explicit using a purely Eulerian formulation.

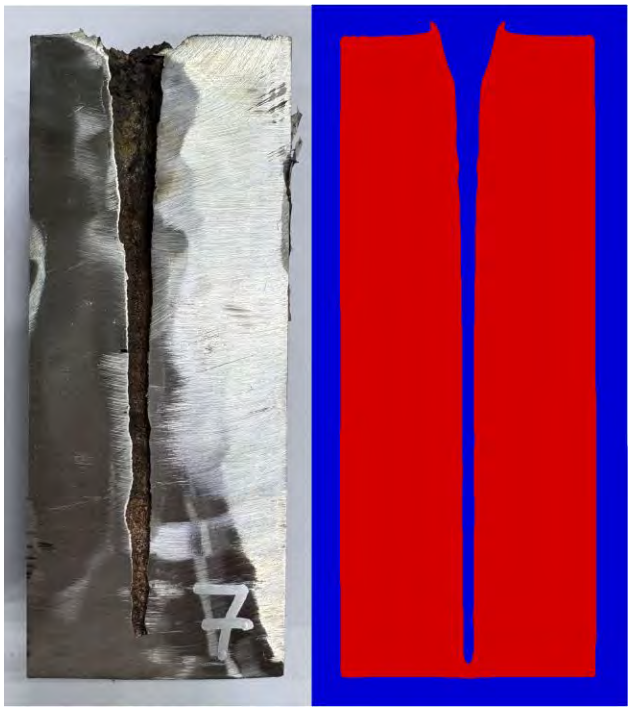


Figure 7. Comparison of experimentally observed (left) and numerically predicted (right) penetration cavity profiles produced in a steel target by an HMX-filled shaped-charge warhead at a standoff distance of 2.6 CD.

Comparison of the experimental observations and numerical predictions demonstrated good agreement in terms of penetration depth, penetration cavity geometry, and entry-hole characteristics. The results indicate that the proposed numerical model is capable of accurately reproducing the essential features of jet formation and penetration, providing a reliable tool for further analysis and optimization of shaped-charge warhead performance.

The present investigation represents the initial stage of an ongoing research effort. Future work will focus on a more comprehensive experimental and numerical assessment of the factors affecting warhead performance, including: (i) the influence of standoff distance on penetration effectiveness, (ii) the effects of liner mass and geometric tolerances, (iii) the influence of explosive charge density, (iv) the effects of case and booster-housing materials, and (v) the influence of initiation conditions on jet formation and overall warhead performance.

The validated numerical model provides a cost-effective framework for the analysis and optimization of shaped-charge warheads, reducing the need for extensive experimental testing during the design process.

Acknowledgements

This research was financially and technically supported by HI Poliex AD, Research and Development Center, Berane, Montenegro. The experimental program was conducted using the company's proving-ground infrastructure, manufacturing facilities, explosive-processing capabilities, technical resources, and expert support. The authors gratefully acknowledge the company's contribution to the preparation of explosive charges, fabrication and assembly of warhead components, safe execution of the experimental tests, and post-test examination of the targets. Special thanks are extended to the technical and engineering staff for their valuable expertise and assistance throughout the experimental campaign.

P.E. acknowledges the financial support provided by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under Contract No. 451-03-34/2026-03/200105, dated 5 February 2026.

References

[1] WALTERS, W.P., ZUKAS, J.A.: *Fundamentals of shaped charges*, John Wiley & Sons, New York, 1989.

[2] ZUKAS, J.A., WALTERS, W.P.: *Explosive effects and applications*, Springer, New York, 2002.

[3] CARLUCCI, D.E., JACOBSON, S.S.: *Ballistics: Theory and design of guns and ammunition*, CRC Press, Boca Raton, FL, 2018.

[4] ELEK, P.: *Terminal ballistics* (in Serbian), University of Belgrade – Faculty of Mechanical Engineering, 2018.

[5] HAZELL, P.J.: Armor: Materials, Theory, Design, CRC Press, 2016.

[6] DSSC, *Abaqus Theory Manual*, Dassault Systemes, Simulia Corp, Providence, RI, USA, 2025.

[7] DEHESTANI, P., FATHI, A., DANIALI, H.M.: Numerical study of the stand-off distance and liner thickness effect on the penetration depth efficiency of shaped charge process, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 233(3) (2019) 977-986.

[8] PYKA, D. et al: Experimental and numerical analysis of piercing with shaped charge of armor steel, Journal of Physics: Conference Series, 2647 (2024), 082017.

[9] ELEK, P., MARKOVIĆ, M., JEVTIĆ, D., ĐUROVIĆ, R.: Modeling of shaped charge jet penetration depth: Analytical and numerical approach, *Scientific Technical Review*, 74 (2) (2024) 17-22.

[10] ELEK, P., MARKOVIĆ, M., JEVTIĆ, D., ĐUROVIĆ, R.: *Modeling of penetration depth of a shaped charge jet*, 9th Congress of the Serbian Society of Mechanics, Vrnjačka Banja, Serbia, July 5-7, 2023.

[11] ELEK, P., MARKOVIC, M., JEVTIC, D., DJUROVIC, R.: *Numerical and analytical modeling of a shaped charge penetration depth*, Proceedings of the 4th International Conference on Impact Loading of Structures and Materials, Freiburg, Germany, 13-17 May 2024, pp. 88-89.

[12] JOHNSON, G.R., COOK, W.H.: *A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures*, Proceedings of the 7th International Symposium on Ballistics, The Hague, 1983.

[13] JOHNSON, G.R., W.H. COOK: Fracture characteristics of three metals subjected to various strains, strain rates, temperatures and pressures, *Engineering Fracture Mechanics*, 21 (1985) 31–48.

[14] PASCU, A.M.; STEFANOAEA, N.: Strain Rate Impact into the Stress and Strain Values at Break of the PA6 GF30-Reinforced Polyamides. *Applied Sciences*, 15, (2025), 11454.

[15] LEE, E., FINGER, M. COLLINS, W: JWL equation state coefficients for high explosives, Lawrence Livermore Laboratory, University of California, 1973.

[16] DOBRATZ, B.M.: *Properties of chemical explosives and explosive simulants*, Lawrence Livermore Laboratory, University of California, 1972.

[17] ELEK, P., DŽINGALAŠEVIĆ, V., JARAMAZ, S., MICKOVIĆ, D.: Determination of detonation products equation of state from cylinder test: Analytical model and numerical analysis, *Thermal Science*, 19(1) (2015) 35-48.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ANALYSIS OF ENERGY DISSIPATION DURING THE IMPACT OF AIR WEAPON PROJECTILES INTO GRANULAR MATERIALS – COMPARISON OF ANALYTICAL MODEL AND EXPERIMENTAL RESULTS

DARKO JANKOVIC
University of Defence, Military Academy, darkoojanko@gmail.com

MARINA SIMOVIC PAVLOVIC
University of Belgrade, Faculty of Mechanical Engineering,, Belgrade, simovicmarina99@gmail.com

NADA CITAKOVIC
University of Defence, Military Academy, Belgrade, nadac@list.ru

LANA MILOSAVLJEVIC
University of Defence, Military Academy, Belgrade, abvgd012@gmail.com

LJUBISA TOMIC
Military Technical Institute Belgrade, ljubisa.tomic@gmail.com

MARTIN OCSENHOFER
University of Vienna, Faculty of Development Studies, martin.ochsenhofer@gmx.at

DARKO VASILJEVIC
Nanophotonics Lab, Photonics Center, Institute of Physics, Belgrade, darkovas@gmail.com

Abstract: *This paper presents an analysis of energy dissipation during the impact of projectiles fired from an air rifle and an air pistol into granular materials of different grain sizes. The study aims to compare analytically calculated energy values with experimental results using an integrated approach that combines mathematical modeling and high-speed imaging. Projectile velocity immediately before impact was estimated using numerical simulations in MATLAB, accounting for aerodynamic drag. Based on this velocity, the projectile's kinetic energy and its distribution into deformation, heat, vibration, and residual energy were determined. The experimental setup included controlled firing conditions with sand targets of varying grain sizes. The impact process was recorded using digital imaging, and the recorded image sequences were analyzed to estimate particle motion, material displacement, and morphological changes in the granular targets. The results indicate a reasonable consistency between the analytical model and experimental estimates for coarse sand. The observed deviations may be associated with granular heterogeneity, projectile geometry, and uncertainties inherent in the optical measurement method.*

Keywords*: energy dissipation, projectile impact, granular materials, kinetic energy, air rifle, air pistol, numerical modeling, high-speed imaging, experimental analysis*

1. INTRODUCTION

Energy dissipation during the impact of a projectile into granular materials represents a complex physical process involving the interaction between the projectile and the discrete structure of the material [1]. This phenomenon has significant applications in ballistics, the development of protective systems, and the analysis of material behavior under dynamic loading [2].

In the case of low kinetic energy projectiles, such as pellets from air weapons, the effects of energy dissipation depend on a large number of parameters: projectile velocity, its geometry, angle of impact, as well as the granulometric composition of the target [3, 4].

The aim of this paper is to compare an analytical model with experimental results of energy dissipation, with a particular focus on the influence of weapon type and material granulation.

2. THEORETICAL AND EXPERIMENTAL METHODOLOGY

In order to achieve a comprehensive understanding of energy dissipation mechanisms, a combined theoretical and experimental approach was adopted. The

Corresponding author: D. Jankovic;
E-mail: darkoojanko@gmail.com
DOI: 10.5937/OTEH26043J

mathematical model provides a basis for predicting projectile behavior, while the experimental methodology enables validation and detailed analysis of the impact phenomena.

2.1. Mathematical Model

The motion of the projectile is modeled as the motion of a material point under the influence of gravity and aerodynamic drag [3, 4]. The kinetic energy of the projectile is defined by the standard expression [3]:

$$E_k = \frac{mv^2}{2} \quad (1)$$

The projectile velocity immediately before impact is determined by numerically solving the differential equations of motion using the Runge–Kutta method [3, 5]. The model includes the following parameters: projectile mass, initial velocity, air resistance coefficient, and air density.

Based on the calculated velocity, the total kinetic energy is determined and further distributed into the following components [3]:

- Energy transferred to the target E_m,
- Plastic deformation energy E_d,
- Thermal energy E_T,
- Sound and shock wave energy E_v
- Residual energy E_R .

The kinetic energy is not entirely transferred to the target, but is divided into several components and is expressed as follows [3]:

$$E_k = E_m + E_d + E_T + E_v + E_R \quad (2)$$

- In the case of granular targets, the energy transferred to the target consists of three main components [3, 5]:
- compression energy — grain separation E_1,
- inertial energy of the grains within the target E_2 ,
- friction energy E_4 .

The compression energy is calculated using the following formula [3]:

$$E_1 = \frac{am}{2b}\ln(1+\frac{b}{a}v_0^2) \quad (3)$$

The inertial component is obtained as the remainder of the total work in the target after subtracting the compression energy and, if applicable, the other energy components [3]. Unlike the compression and inertial components, the friction energy is not explicitly separated in the basic Poncelet formulation, but is introduced in practice as an empirical approximation [3]:

$$E_4 = E_m f_0(1+\alpha\sin^2\theta) \quad (4)$$

with typical values of approximately $f_0 \approx 0.05$, i.e. 5% for normal impact, and α ≈ 1 for oblique impacts. This approximation is based on measurement results and experimental calibrations. It represents an empirical approximation based on measurements [3].

In terminal ballistics, the angle θ denotes the angle at which the projectile impacts the target relative to the normal to its surface.

This distribution serves as the basis for comparison with the calculated and experimentally determined energy values presented in the Results and Analysis section.

2.2. Experimental Methodology

The experimental investigation was conducted using two types of air weapons: the STEYR LG 110 air rifle and the STEYR LP 10 air pistol. Their relevant technical characteristics are presented in Table 1 [6, 7].

Table 1. Technical characteristics of the STEYR LG 110 air rifle and the STEYR LP 10 air pistol [6, 7]

Characteristics	STEYR LG 110 (Rifle)	STEYR LP 10 (Pistol)
Overall length	1170 mm	400 mm
Overall height	200 mm	142 mm
Barrel length	550 mm	227 mm
Caliber	4.5 mm / .177 in	4.5 mm / .177 in
Mass	4750 g	968 g
Width	—	50 mm
Propulsion	Compressed air	Compressed air
Trigger	Mechanical	Mechanical
Trigger pull	50–250 g	50–250 g
Muzzle energy	24 J	17 J
Initial projectile velocity	240 m/s	160 m/s

Furthermore, the experimental conditions were defined through a combination of parameters such as the firing angle, projectile type, and target granulation [5]. The targets were composed of sand with three different granulations, whose characteristics are shown in Table 2 [5].

Table 2. Characteristics of granular materials [5]

Sand Type	Specific density [kg/m3]	Granulation [mm]
Coarse sand	1770	~4-10
Medium-grained sand	1700	~3
Fine sand	1340	~0-0.3

For recording and analyzing the impact process, a Panasonic DMC-FZ200 digital camera was used at a recording rate of 40 frames per second [8].

For the analysis of the granular target response after projectile impact, an existing C++-based image-

processing program using the OpenCV library was modified and adapted to the specific requirements of the present experimental investigation [5].

The image-processing procedure involved tracking selected reference particles in consecutive frames recorded at 40 frames per second, corresponding to a time interval of 0.025 s. Particle velocities were estimated from their displacements between successive frames. Due to particle overlap, occlusion, and motion blur, only clearly identifiable reference particles were included in the tracking procedure.

Image segmentation and contour analysis were applied to characterize morphological changes in the granular target, including particle cloud formation, crater geometry, and subsequent erosion. The granular targets were placed in wooden frames measuring 640 × 640 × 120 mm (width × height × depth), providing known geometric dimensions of the experimental setup. Geometric corrections were considered to account for the inclination of the target and the angular position of the camera.

The resulting measurements were used to estimate particle velocities and the spatial characteristics of displaced granular material. The analysis was subject to limitations associated with image resolution, particle overlap, and the estimation of three-dimensional quantities from two-dimensional images.

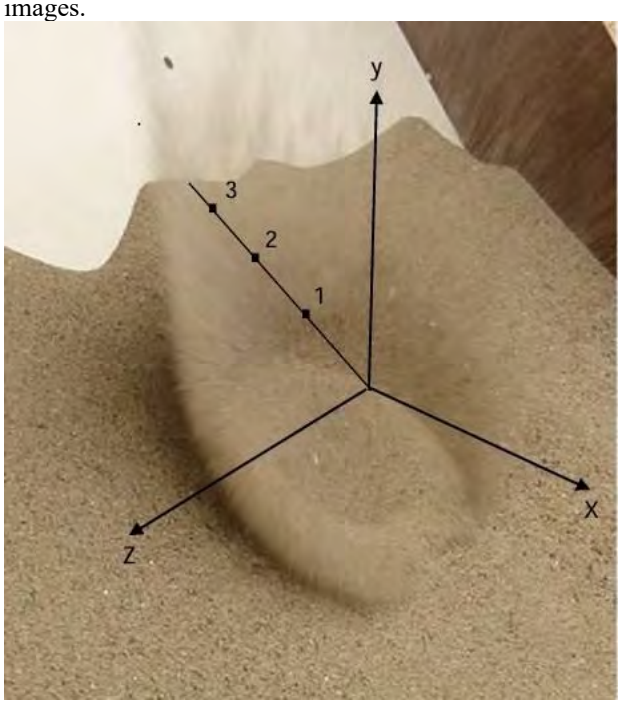


Figure 1. Analysis of photographs of the sand surface

3. RESULTS AND ANALYSIS

Analytical calculations of projectile velocity and energy, as well as their distribution, are presented in Table 3.

Table 3 – Comparison of calculated compression energy and experimentally determined energy

Air weapon type	Sand Type	Calculated compression energy E1 [J]	Experimental energy Eexp [J]
STEYR LG 110	Coarse sand	0.566	0.618
STEYR LP 10	Coarse sand	0.470	0.395

The experimental energy values are estimates based on the volume of displaced material, its density, and the average velocity of selected reference particles. The quantitative comparison was limited to coarse-grained sand due to the greater reliability of optical particle velocity measurements.

Based on the data from Table 3, it can be observed that the projectile fired from the air rifle produces higher calculated compression energy and higher experimentally determined energy compared to the projectile fired from the air pistol, which directly affects the intensity of energy dissipation in the target [5].

Each experimental configuration was repeated three times (n = 3). For coarse sand under normal impact conditions, the standard deviation of the recorded particle cloud volume was 6,844.57 mm³ for the STEYR LG 110 rifle and 1,550.20 mm³ for the STEYR LP 10 pistol. These values characterize the variability of the volumetric response between repeated experiments. The maximum recorded volume was used for the subsequent energy estimation.

In addition to the comparison between analytically calculated compression energy and experimentally estimated energy presented in Table 3, the distribution of the projectile kinetic energy into individual components was also analyzed. The calculated energy values for both air weapons impacting coarse sand at an angle of 0° are presented in Table 4.

Table 4 - Model-based energy components for projectile impact into coarse sand at 0°

Energy component	LG 110 [J]	LP 10 [J]
Compression energy	0.566	0.470
Inertial energy	9.721	4.253
Friction energy	0.541	0.248
Deformation energy	0.343	0.158
Thermal energy	0.162	0.075
Vibration energy	0.114	0.053
Residual energy	0.000	0.000

As shown in Table 4, the inertial component accounts for the largest proportion of the projectile kinetic energy, while smaller contributions are associated with compression, friction, deformation, thermal effects, and vibrations. These components were calculated using adopted coefficients of the mathematical model. Thermal and vibration-related energies were not directly measured, and no temperature measurements were performed during the experiments. Therefore, the presented values represent a model-based energy distribution rather than independently measured energy components.

The analysis of the experimental results shows significant differences in material behavior depending on granulation [5]. The cumulative volumetric particle size distribution of the three sand types, presented in Figure 2, shows clear differences in the internal structure of fine sand, medium-grained sand, and coarse sand [5]. These differences are important for understanding the mechanisms of energy dissipation in granular materials.

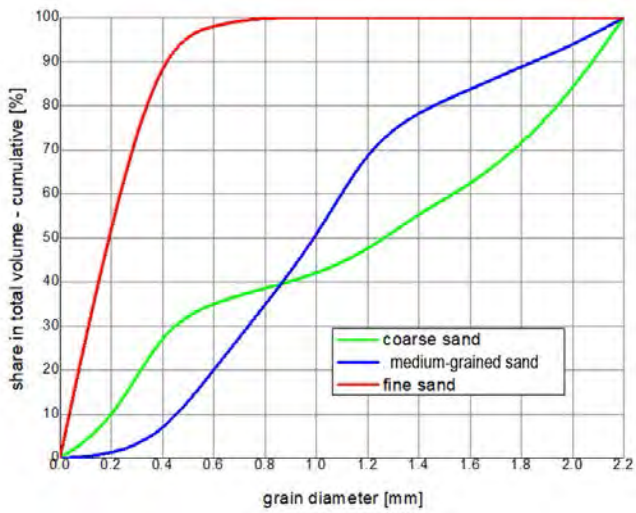


Figure 2. Cumulative volumetric particle size distribution as a function of particle diameter for three sand types

In the case of fine sand, localized energy dissipation is observed, manifested through a small "cloud" volume and limited material displacement. In contrast, coarse sand exhibits more pronounced material displacement and a more gradual dissipation of energy in deeper layers [5]. In this analysis, the combined volume of the particle "cloud" and the locally displaced material in the impact zone is treated as the total impact-related displaced volume. This behavior is illustrated in Figure 3, where the combined cloud and impact volume shows a significantly stronger response of coarse sand compared to fine sand [5].

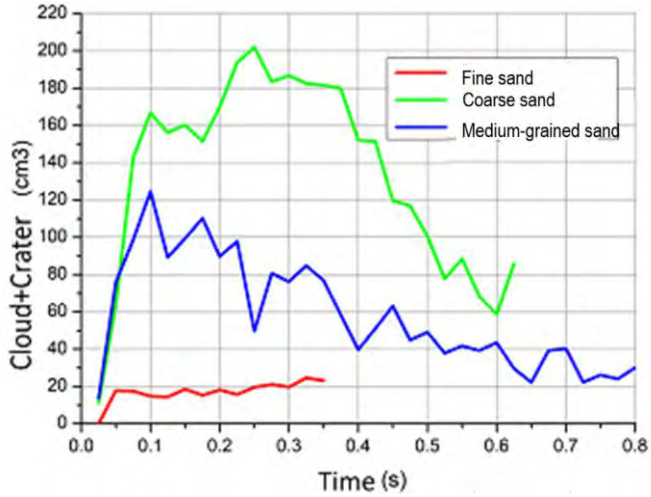


Figure 3. Change in combined cloud and impact-related displaced volume as a function of time for an air rifle shot using a flat-tip diabolo projectile at a firing angle of 0°

The differences between the analytically calculated and experimentally estimated energy values can be partly attributed to the heterogeneous structure of granular materials and the geometry of the projectile. Unlike a homogeneous continuous medium, sand consists of discrete particles with different sizes, shapes, and local packing arrangements. During impact, these characteristics affect interparticle contacts, friction, grain rearrangement, and the spatial distribution of transferred energy. Consequently, the actual granular response may differ from the simplified resistance assumptions adopted in the analytical model.

Projectile tip geometry also influences the interaction between the projectile and the granular target. Flat-tip projectiles produce a more pronounced initial displacement of grains, while conical-tip projectiles exhibit a different spatial and temporal distribution of material displacement and erosion. These differences are consistent with the experimentally observed variations in particle cloud formation and erosion dynamics.

Additional discrepancies may arise from the limitations of optical particle tracking, particularly particle overlap, irregular trajectories, and the use of average particle velocities in estimating the kinetic energy of displaced material. Therefore, the observed differences should be interpreted as the combined effects of granular heterogeneity, projectile geometry, and measurement uncertainty, rather than as the consequence of a single isolated factor.

The obtained results also indicate that the shape of the projectile tip significantly affects the mode of energy transfer [5]. This is consistent with previous analyses of pellet geometry and aerodynamic behavior [9, 10]. The experimental observations indicate that flat-tip projectiles can produce higher and more irregular material displacement and erosion, whereas conical-tip projectiles exhibit a different, more time-distributed granular response [5].

4. CONCLUSION

The results indicate reasonable consistency between the analytical model and the experimentally estimated energy values for coarse sand. The observed differences reflect the complexity of granular material behavior and the limitations of the optical measurement method. Further experimental investigations and statistical analysis are required to establish the model's predictive accuracy across different granular materials.

The obtained results may be applied in the analysis of protective materials and in the optimization of structures exposed to impact loading.

References

[1] HERRMANN,H.J.: Granular matter, Physica A: Statistical Mechanics and its Applications, 313(1–2), 2002, 188–210.

[2] ZHANG,Y., ZHU,F., LU,G., WANG,Z.: Ballistic performance of protective structures using sand and foam-filled layers, International Journal of Heat and Mass Transfer, 161, 2020, 120268.

[3] CARLUCCI,D.E., JACOBSON,S.S.: Ballistics: Theory and Design of Guns and Ammunition, 3rd ed., CRC Press, Boca Raton, 2018.

[4] ANDERSON,J.D.: Fundamentals of Aerodynamics, 5th ed., McGraw-Hill Education, New York, 2010.

[5] JANKOVIĆ,D.D., PAVLOVIĆ,S.M., NIKOLIĆ,Z.M., TOMIĆ,L.D., JERKOVIĆ,D.D., VASILJEVIĆ,D.M.: Using image analysis to determine the energy transfer profile of a projectile impact in a granular substrate, Science of Sintering, 00(00), 2025, 24.

[6] STEYR SPORT GMBH: Operating Manual – Steyr LG 110 Match / Light / Bench Rest / Rapid Fire, Steyr Sport GmbH, Raming, Austria, 2023.

[7] STEYR SPORT GMBH: Operating Manual – Steyr LP 10 / Compact, Steyr Sport GmbH, Raming, Austria, 2023.

[8] PANASONIC: Lumix DMC-FZ200 – Digital Camera Specification, Panasonic Global.

[9] LADOMMATOS,N.: Drag coefficients of air rifle pellets with wide range of geometries, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 235(23), 2021, 6133–6152.

[10] RAFEIE,M., TEYMOURTASH,A.R.: Aerodynamic and dynamic analyses of three common 4.5 mm-caliber pellets in a transonic flow, Scientia Iranica, Transactions B: Mechanical Engineering, 23(4), 2016, 1767–1776.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SAPONIFICATION AS AN ENABLING STEP FOR FAST GC DETERMINATION OF STABILIZERS IN SMOKELESS POWDERS

TEODORA STANČIĆ
Military Technical Institute, Belgrade, teodorastancic@gmail.com

NADA MARINKOVIĆ
Milan Blagojević Namenska, Lučani, nadam966@gmail.com

SUZANA KOVAČEVIĆ
Milan Blagojević Namenska, Lučani, skovacevic135@gmail.com

MIRJANA MILEKIĆ
Milan Blagojević Namenska, Lučani, mirjanamilekic84@gmail.com

VESNA PETROVIĆ
Technical Overhaul Facility Kragujevac, lab@trzk.co.rs

MIRJANA KRSTOVIĆ
Military technical institute, Belgrade, mirjana.krstovic@mod.gov.rs

Abstract: *Smokeless powders represent chemically unstable systems whose safety and performance critically depend on the integrity of their non-energetic organic components. Their primary constituent, nitrocellulose, is inherently prone to thermal decomposition, which under unfavorable or long-term storage conditions may lead to autocatalytic degradation and, ultimately, self-ignition of ammunition. In this context, fast and reliable determination of organic stabilizers and additives by gas chromatography (GC) is not only an analytical requirement but a safety imperative in both production and storage environments. In this work, saponification is examined as a key step in sample preparation with the potential to significantly accelerate the analytical workflow. Its performance was directly compared with conventional preparation procedures that do not include saponification, specifically extraction in methanol and extraction in dichloromethane. These approaches were systematically matched under equivalent conditions prior to GC analysis to clearly assess the impact of the saponification step on both efficiency and analytical performance. All investigated procedures provided reliable and reproducible quantification of the non-energetic organic components present in the analyzed samples by GC. However, the saponification-based method enabled a substantial reduction in total sample preparation time, by several hours, while maintaining analytical accuracy and precision. These findings demonstrate that the introduction of saponification as an initial preparation step offers a practical route toward faster GC analysis without compromising data quality. This improvement is particularly relevant for routine quality control, where reduced turnaround time directly enhances process monitoring and operational efficiency.*

Keywords: *smokeless powders, gas chromatography (GC), sample preparation, stabilizers, saponification.*

1. INTRODUCTION

Although the field of composite propulsion materials is developing rapidly, gunpowder and propellants based on nitrocellulose remain the pillar of ammunition technology in small caliber and standard artillery ammunition. Due to the inherent susceptibility of nitrocellulose to chemical degradation [1-4], extensive research has established various instrumental and classical methods to monitor its chemical aging as quality control procedures [5]. Consequently, precise criteria have been defined for evaluating and categorizing the stability of gunpowder and rocket propellants over time. However, before assessing long-term chemical stability, a critical phase in the ammunition life cycle is rigorous quality control during both production and product-acceptance phases. A key aspect of this process is the precise quantification of essential organic components, especially stabilizers and plasticizers. To meet the rigorous demands of quantifying these critical parameters, chromatographic methods emerge as the premier analytical tool, enabling rapid, efficient, and simultaneous identification and quantification of organic compounds within the propellant matrix. [6]

Therefore, the development and implementation of rapid, reliable, and precise analytical methods is essential for manufacturing facilities dedicated to smokeless powder production. While standard chromatographic protocols provide highly accurate results within brief timeframes, sample throughput remains limited by the pre-analytical

Corresponding author: T.Stančić;
E-mail: teodorastancic@gmail.com
DOI: 10.5937/OTEH26044S

preparation phase. Gas chromatography (GC) generally requires between 5 and 10 minutes for analysis. Consequently, the primary obstacle in achieving rapid and reliable analytical systems shifts from instrumental analysis to sample preparation, which demands optimization for exhaustive extraction within the shortest possible time.

International [1] and domestic standards [7-8] aim to establish versatile testing methodologies. These protocols must accommodate an exceptionally wide range of propellant grain geometries, dimensions, and shapes, while seamlessly adapting to the diverse chemical compositions found in modern formulations.

The primary objective of this research is to develop a rapid sample preparation method for production-line gunpowder samples prior to gas chromatographic testing. Given that the exact chemical composition of the control samples was predetermined, three preparation techniques were employed in parallel. The study aims to achieve the fastest possible analytical workflow time without compromising the accuracy and sensitivity. For this reason, the role of saponification in the sample preparation was rigorously evaluated, focusing specifically on its impact on overall analysis speed and extraction efficiency.

2. MATERIALS AND METHODS

The nitrocellulose-based smokeless gunpowder samples used in this research were produced within the last six months at "Milan Blagojević – Namenska" factory.

These samples vary in shape and size and were specifically chosen to demonstrate the diversity of nitrocellulose powders, not only in terms of composition and application but also regarding grain geometry and dimensions. As can be seen in Figure 1, nitrocellulose-based gunpowder can take on various forms depending on the desired burning rate and the specific purpose of the system it serves. Because these physical properties significantly influence ballistic performance, the development of a universal method applicable under identical conditions to all types of gunpowder and rocket fuel is essential. Establishing such a standard aims to unify methodology across the field, ensuring that analytical results remain consistent and comparable regardless of the propellant's specific geometry and compositions.

Figure 1. from left to right: samples of NGB 261, NC-27 and SB 715

Information regarding the selected non-energetic organic compounds (diphenylamine – DPA, ethyl centralite -EC, dibutyl phthalate – DBP, camphor – camp, triethyl acetate -TAC) of the ten selected powders is presented in Table 1. However, since all compositions are currently in operational use, the disclosure of complete formulations or the mass fractions of individual components is restricted.

Table 1. Selected samples of gunpowder

Sample name	Non-energetic organic compounds				
	DPA	EC	DBP	Camp	TAC
NC-46	+			+	
NC-27	+				
NC-49	+				
NC-261	+		+		
SB 620	+	+			
SB 511	+		+		
SB 411	+				
SB 715	+		+		
NGB 051		+			+
NGB 161		+			+

2.1. Gas chromatography

Gas chromatography with a capillary column works by vaporizing a sample and carrying it through a thin, long column using an inert carrier gas. Inside the column, components are separated based on their unique chemical affinity for the stationary phase coating the internal walls. Each gunpowder organic component exits the column at a specific retention time, allowing the detector to identify and quantify them individually.

In this research, for separation, identification and quantification of components of interests, Agilent gas chromatograph model 7890 with flame ionization detector (FID), with autosampler, was used. Capillary column HP-5 has been employed as separating tool. The chromatographic conditions were as follows:

- Injector temperature: 130 °C,
- Initial oven temperature: 130 °C,
- heating rate of 10 °C/min,
- final temperature of 210 °C,
- Detector temperature: 300 °C,
- Carrier gas: Hydrogen (H_2),
- Hydrogen flow rate: 25 mL/min,
- Air flow rate: 400 mL/min.

While high-performance liquid chromatography (HPLC) [9-12] is the recommended method for analyzing older powders, due to the tendency of diphenylamine (DPA) derivatives to convert back into DPA at the elevated temperatures required by GC, creating a false representation of the active stabilizer content, GC remains a highly favorable method for the quality control of new powders because economic and ecologic efficiency. In newly manufactured powders, the DPA has not yet had time to degrade into its derivatives, therefore, the risk of thermal reversion is non-existent, and the results remain accurate. Furthermore, GC offers several technical advantages for these samples: superior column resolution, high precision separation of other components in

gunpowder (especially if sample contain nitroglycerin (NG), dinitrotoluene (DNT) and trinitrotoluene (TNT).

3. EXPERIMENTAL PART

All ten selected powders were prepared using three different methods, with each test performed in duplicate. Since the powders are intended for ammunition up to 20 mm and possess a small original grain size, no preliminary grinding of the samples was required.

3.1. Sample preparation in methanol

A sample of gunpowder 1.000 ±0.001 g was weighed and transferred into a 100 cm^3 volumetric flask. Methanol was added to half of the flask's volume, and the mixture was left for 4-6 hours. Subsequently, the flask was filled to the mark with methanol and agitated on a magnetic stirrer for 6 to 8 hours. The sample requires centrifugation to facilitate nitrocellulose precipitation before proceeding with further analytical steps. A 20 cm^3 volume of the prepared solution was pipetted into a 50 cm^3 volumetric flask, followed by the addition of 5 cm^3 of the internal standard solution (1.5 g dimethyl phthalate dissolved in 250 cm^3 of methanol), and then diluted to the mark with methanol. The sample was then transferred to a vial for autosampler, as it can be seen in Figure 2. To obtain satisfactory results, it is essential to pass the solution through a 0.5 µm PTFE filter prior to usage.

Since an internal standard was introduced into the sample, the concentration of the analyzed substances is calculated using the following equation (1):

$$A = \frac{P_A}{m_u * P_{IS}} * k_A \qquad (1)$$

Where are:

- k_A - mean value of two replicates of proportionality constant for the analyzed component, calculated by instruction from [8]
- m_u - sample mass (g),
- P_A - peak area of the analyzed component (%),
- A - content of the analyzed component in the gunpowder sample (%),
- P_{IS} – peak area of the internal standard (%).

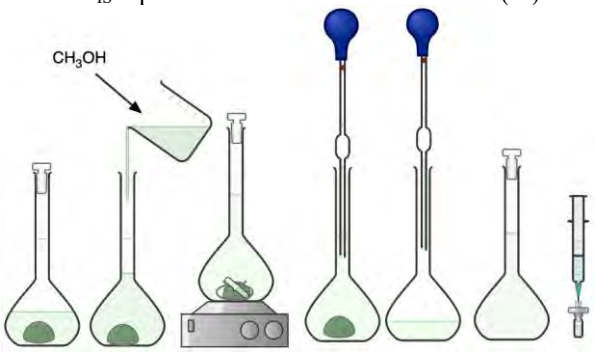


Figure 2. Sample preparation in methanol, from left to right: sample with ½ volume of methanol for 4-6 h, adding methanol to full volume into the flask, stirring sample for 6-8 h, 20 cm^3 volume pipetted and transferred into the new flask, followed by the addition of 5 cm^3 of the internal standard solution, filtrating sample and transferring it to vial

In a 50 cm^3 volumetric flask, weigh the required amount of the component to be determined for calibration and add 25 cm^3 of solvent. Subsequently, add 5 cm^3 of the previously prepared internal standard solution, then dilute to the mark with the solvent. Calibration is performed using standard solutions by testing the organic components under the same chromatographic conditions used for determining organic additives in gunpowder samples.

The calculation of the proportionality constant is performed according to the following equation (2):

$$k_A = \frac{m_A * P_{IS}}{P_A} * 100 \qquad (2)$$

Where are:

- k_A - proportionality constant for the analyzed component;
- m_A - mass of the component weighed for the preparation of calibration solutions (g);
- P_A - peak area (signal) of the analyzed component (%);
- P_{IS} - peak area of the internal standard (%).

3.2. Sample preparation in dichloromethane

A sample of gunpowder 1.000 ±0.001 g was weighed and transferred into a 50 cm^3 volumetric flask. Dichloromethane was added to half of the flask's volume, and the mixture was left for 24 hours, as it can be seen in Figure 3, step by step.

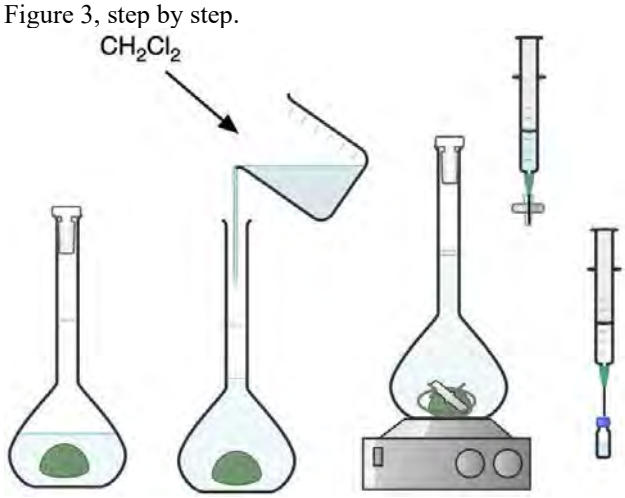


Figure 3. Sample preparation in dichloromethane, from left to right: sample with ½ volume of dichloromethane 24h left in dark place, introducing 5 cm^3 internal standard and adding dichloromethane to full volume into the flask, stirring sample for 4-6 h, filtrating sample and transferring it to vial

Subsequently, the in flask was added 5 cm^3 of the internal standard solution (1.5 g dimethyl phthalate dissolved in 250 cm^3 of dichloromethane) and filled the flask to the mark with dichloromethane. Sample prepared into that manner was agitated on a magnetic stirrer for 4 to 6 hours. The undiluted sample was then transferred to a vial for autosampler. To obtain satisfactory results, it is essential to pass the solution through a 0.5 µm PTFE filter prior to usage. With the internal standard incorporated, the

calculation of target component concentrations is performed using equations (1) and (2) from the section 3.1.

3.3. Sample preparation by saponification

Approximately 5 g of gunpowder was weighed into a flat-bottomed flask, followed by the addition of 80 cm^3 of a 20% (m/m) NaOH solution. The apparatus was assembled for reflux using a reflux condenser. The flask was placed on a heating mantle and refluxed for approximately 2 hours. Without removing the condenser, the flask was cooled in a water bath to room temperature. The condenser was then rinsed with 15 to 25 cm^3 of dichloromethane (Figure 4), and the contents of the flask were transferred to a laboratory separatory funnel, which was firstly added 200 cm^3 of a 5% (m/m) NaCl solution.

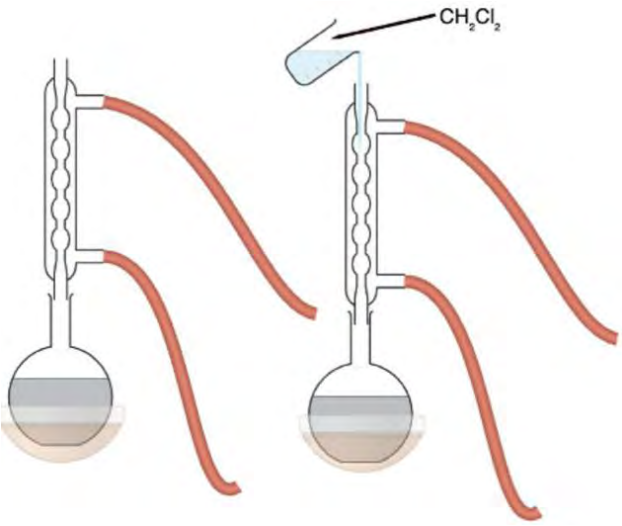


Figure 4. Heating solution of gunpowder and NaOH with reflux for 2h, colling down to the room temperature and rinsed by dichloromethane

The content in separatory funnel was then stoppered and shaken vigorously. The lower layer from the funnel was transferred into a 200 cm^3 volumetric flask. The extraction process from the flask and the separatory funnel was repeated twice more with dichloromethane to ensure quantitative transfer is done. The volumetric flask was then filled to the mark with dichloromethane, as it can be seen in Figure 5.

The sample prepared in this manner was used for the determination of organic components via gas chromatography as follows: 5 cm^3 of the internal standard was added to a 50 cm^3 volumetric flask, which was then filled to the mark with the previously prepared sample. The prepared sample was transferred without dilution into a vial, from which it was injected into the chromatographic column via autosampler or manual injection. At least two replicates were prepared and analyzed for each gunpowder sample.

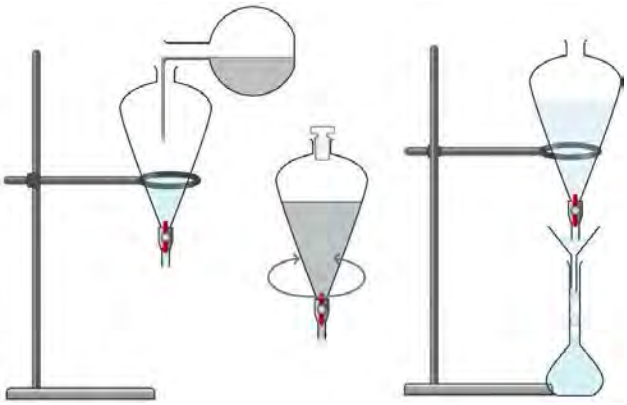


Figure 5. Separation process by using separatory funnel

For the quantification of the target components, it is necessary to prepare calibration solutions for all analyzed components within an appropriate concentration range. Calibration is performed under the same chromatographic conditions as those used for sample testing.

4.RESULTS AND DISCUSION

The stabilizer and other non-energetic organic components content for four different singlebase gunpowder, for all three sample preparation technique, are presented in Table 2. The results were reported as mass percentages, scaled to two decimal places.

Table 2. Results of measuring of other non-energetic organic compound in selected single-based gunpowder

Sample name	Non-energetic organic compounds			
	sample prep.	DPA	DBP	Camp
NC-46	dichloromethane	1.30	/	1.34
	methanol	1.32	/	1.36
	saponification	1.26	/	1.29
NC-27	dichloromethane	1.53	/	/
	methanol	1.50	/	/
	saponification	1.52	/	/
NC-49	dichloromethane	1.33	/	/
	methanol	1.30	/	/
	saponification	1.28	/	/
NC-261	dichloromethane	1.46	3.36	/
	methanol	1.41	3.41	/
	saponification	1.42	/	/

As it can be seen, the relative error between different methods of sample preparation ranges from 1.9 to 2.1%. Furthermore, no consistent trend was observed to indicate that a specific preparation method systematically yields higher values across all sample types. Additionally, a notable deviation was detected in the results for each evaluated component across all three preparation protocols. Nevertheless, all observed values remain within an acceptable margin of error.

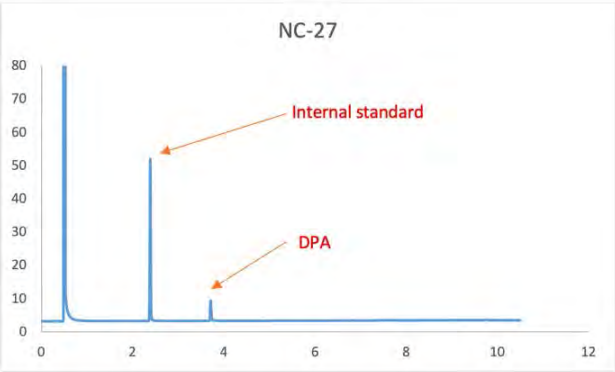


Figure 6. Chromatograms of single-base gunpowder NC-27, extraction in methanol

Regarding spherical (ball) gunpowder, the results for the determination of the content of non-energetic compounds, obtained by analyzing three selected compositions and employing three different sample preparation techniques, are presented in Table 3.

Table 3. Results of measuring of non-energetic organic compound in selected spherical (ball) gunpowder

Sample name	Non-energetic organic compounds			
	sample prep.	DPA	DBP	EC
SB-511	dichloromethane	1.33	5.30	/
	methanol	1.32	5.21	/
	saponification	1.31	/	/
SB-411	dichloromethane	1.37	/	1.32
	methanol	1.36	/	1.31
	saponification	1.39	/	1.30
SB-715	dichloromethane	1.18	6.61	/
	methanol	1.20	6.52	/
	saponification	1.18	/	/

For spherical powders, the determination of non-energetic organic components yields significantly better reproducibility. Specifically, the relative error is approximately 1%, and often lower, for the quantification of stabilizer and plasticizer contents across all three sample preparation techniques evaluated for GC analysis.

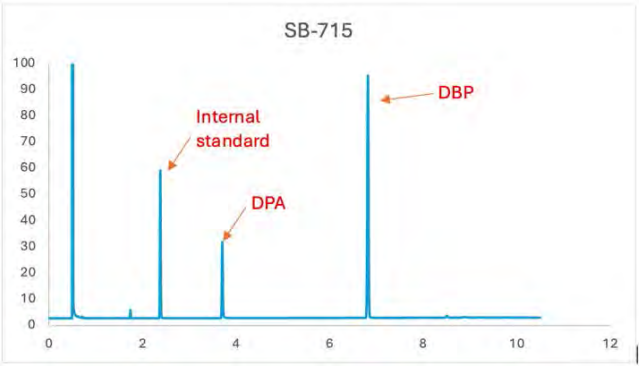


Figure 7. Chromatograms of spherical (ball) gunpowder SB-715, extraction in dichloromethane

Table 4. Results of measuring of non-energetic organic compound in selected double-based gunpowder

Sample name	Non-energetic organic compounds		
	sample prep.	EC	TAC
NGB-051	dichloromethane	1.79	0.40
	methanol	1.76	0.40
	saponification	1.79	/
NGB-161	dichloromethane	1.74	0.29
	methanol	1.79	0.28
	saponification	1.75	/

The results presented in Table 4 indicate nearly ideal agreement between the saponification and dichloromethane preparation protocols for double-base gunpowder. In contrast, the methanol-based preparation yields variations without a consistent trend, the values obtained by methanol preparation are higher for one sample but lower for another. For the stabilizer content, the relative error ranges between 1% and 1.5% when evaluating the results across all three sample preparation technique.

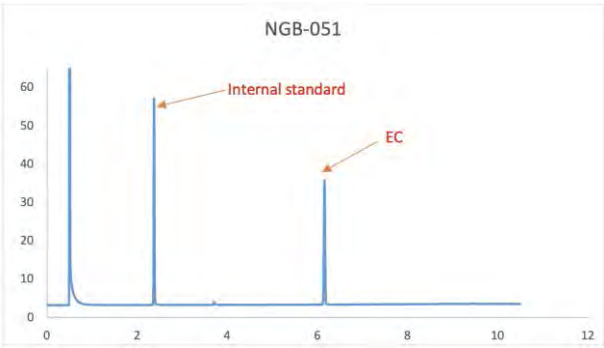


Figure 8. Chromatograms of double-base gunpowder NGB-051, preparation by saponification

5. CONCLUSION

This study focused on evaluating diverse sample preparation methods for gas chromatographic analysis to determine primaly stabilizer and secondary, the other non-energetic compounds in propellants, which serve as essential intermediate and final quality control parameters in manufacturing plants. Based on the findings, every evaluated preparation protocol provides highly reproducible data. The maximum relative error was observed in single-base powders, though it strictly remained below 2.5%, while for all other propellant types, this value was approximately 1% or lower. The primary advantage of the saponification method lies in its rapid sample processing time, less then few hours, compared the other two which demand more between 14 and 30 hours. However, it lacks the capability to determine DBP and triacetin (TAC). Consequently, saponification represents the method of choice when rapid turnaround times are required to obtain immediate feedback on stabilizer concentrations. These stabilizers are critical factors for propellant stability and constitute the primary checkpoint during in-process monitoring, final product

release, and throughout the entire propellant lifecycle. Nevertheless, if the analytical objective requires the comprehensive determination of all non-energetic compounds within the formulation, preparation via methanol or dichloromethane extraction becomes the preferred approach.

Although sample preparation for HPLC is less time-consuming, GC offers a far higher justification, thus, the results of this research carry significant value, as they provide critical insights into the capabilities of various preparation techniques to deliver reliable results through a method that, while perhaps not time-efficient, remains fundamentally sound from both an economic and ecological perspective. The utilization of GC offers distinct advantages, particularly in high-throughput industrial settings such as gunpowder factories, where numerous routine samples are analyzed daily. Economically, GC significantly minimizes operational expenses by employing inert carrier gases instead of the expensive, high-purity solvents required as mobile phases in HPLC. Furthermore, it mitigates the substantial costs associated with high-pressure pump maintenance and column degradation caused by aggressive solvent. From an environmental standpoint, GC operation is fundamentally cleaner and more sustainable, as it virtually eliminates the generation of hazardous liquid chemical waste, thereby bypassing the stringent and costly disposal regulations inherent to HPLC operations. Consequently, for routine quality control in manufacturing plants with heavy daily sample loads, GC represents a highly cost-effective and ecologically viable analytical framework.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325, and University of Defense, Military Academy, Proj. No. VA-TT/1/22-24.

References

[1] AOP-48 Ed.2. *Explosives, nitrocellulose-based propellants - stability test procedures and requirements using stabilizer depletion*, Brussels. North Atlantic Treaty Organization, Military Agency for Standardization, 2008.

[2] RAKIĆ M, KRSTOVIĆ M, MIJATOV S, DIMIĆ M, FIDANOVSKI B. *Comparative analysis of volumetric and instrumental method for determination of nitroglycerin content in gunpowder*. Scientific Technical Review, 2021, 2, pp.41-45. doi: 10.5937/str2102041R

[3] LINDBLOM T. *Reactions in the System Nitro-cellulose/Diphenylamine with Special Reference to the Formation of a Stabilizing Product Bonded to Nitro-cellulose*, Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, 2004. ISBN 91-554-5868-8

[4] KRSTOVIĆ M, RAKIC M, JELISAVAC LJ, DIMIĆ M, MOJSILOVIĆ J, FIDANOVSKI B. *Comparative analysis for determination of stabilizer content in gunpowder and double base propellant by HPLC method.* In: Proceeding of 10th International Scientific Conference on Defense Technologies, OTEH 2022, Belgrade, Serbia, 2022, pp 138-143, ISBN 978-86-81123-85-0 86

[5] BHON M. *Principles of Ageing of Double Base Propellants and its Assessment by Several Methods Following Propellant Properties.* STO-MP-AVT-268

[6] MATEČIĆ MUŠANIĆ S., SUĆESKA M., ČULJAK R. *The Applicability of Chromatographic Methods in the Investigation of Ageing Processes in Double Base Rocket Propellants.* Central European Journal of Energetic Materials, 2013, 10(2), 245-262 ISSN 1733-7178

[7] SORS 8069/91, *Praćenje hemijske stabilnosti baruta i raketnih goriva.* 1991. Beograd.

[8] SORS 1008/85, *Metode analitičkih ispitivanja malodimnih baruta,* 1985. Beograd.

[9] SEROL M., AHMAD S. M., QUINTAS A., FAMILIA C., *Chemical Analysis of Gunpowder and Gunshot Residues*, Molecules 2023, 28(14), 5550; https://doi.org/10.3390/molecules28145550

[10] LÓPEZ-ĹOPEZ, M.; BRAVO, J.C.; GARCIÁ-RUIZ, C.; TORRE, M. *Diphenylamine and Derivatives as Predictors of Gunpowder Age by Means of HPLC and Statistical Models.* Talanta 2013, 103, 214–220.

[11] ŠULTANS J., STOJILJKOVIĆ S., PETROVIĆ V., *Influence of gunpowders and rocket propellants sample preparation on test results*, In: Proceeding of 11th International Scientific Conference on Defense Technologies, OTEH 2024, Tara, Serbia, 2024, pp 252/257 ISBN 978-86-81123-94-2

[12] KRSTOVIĆ M., RAKIĆ M., MIJATOV S., MOJSILOVIĆ J.,ALEKSIĆ S., FIDANOVSKI B. *Evaluating the Effectiveness of Different Sample Preparation Routes for Organic Components Quantification in Gunpowder and Double Base Rocket Propellants for HPLC Analysis*, Scientific Technical Review, Vol. 75, No.1 (2025), pp. 3-10, doi: 10.5937/str2500001K

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ANALYTICAL DETERMINATION OF MEAN BALLISTIC PRESSURE FROM THE ENERGY BALANCE OF PROPELLANT COMBUSTION PRODUCTS WITH EXPERIMENTAL VALIDATION

SLOBODAN ORAŠANIN
Optical Systems DOO, Beograd, orsasniper@gmail.com

MIODRAG LISOV
Pneumatik Flex PPT Cilindri, Nova Pazova, miskolisov@gmail.com

ALEKSA ANIČIĆ
Agency for Testing, Stamping and Marking of Weapons, Devices and Ammunition, Kragujevac, tehnika@aiz.rs

MILOŠ JONČIĆ
Ministry of Foreign Affars, Arms Control department Belgrade, milos.joncic@mfa.rs

MILJKO PAVLOVIĆ
Milan Blagojević – Namenska A.D., Lučani, mpavlovic.mbn@gmail.com

Abstract: *Determining the mean ballistic pressure (MBP) of the propellant gases in small-caliber weapon barrels is essential for the analysis and design of weapon systems. This paper presents a method for its estimation based on the energy balance of the propellant charge, utilizing established internal ballistics models.*
A simple analytical method is proposed under the assumptions of an adiabatic process, ideal gas behavior, and the quasi-stationary flow hypothesis. Based on these assumptions, standard internal ballistics relations were applied to determine the parameters required to calculate the mean ballistic pressure at the moment of projectile muzzle exit, establishing its relationship with the internal energy of the propellant charge.
Experimental validation was performed using piezoelectric pressure sensors mounted on a test barrel. The measured values of the propellant gas pressure were compared with the calculated results of the mean ballistic pressure. The agreement between them was analyzed to evaluate the accuracy and reliability of the proposed method. The results show that the method provides a practical and sufficiently accurate tool for estimating the mean ballistic pressure (MBP) of propellant gases in engineering applications.

Keywords: *mean ballistic pressure, propellant gas energy, interior balistics, Resal's equation, Piobert's Formula*

1. INTRODUCTION

For small arms designers who develop muzzle devices, a necessary data point is the value of the mean ballistic pressure (MBP) at the moment the projectile leaves the barrel. An interior ballistic (IB) calculation provides this data. A problem arises when not all required propellant characteristics, as well as the design features of the cartridge and the weapon necessary for a reliable IB calculation, are known.

Therefore, a need exists in practice for a simplified calculation method that does not require detailed IB data to calculate the MBP. For this method, analytical expressions were developed using the known laws of interior ballistics under certain assumptions and corresponding simplifications. These formulas allow the calculation of the MBP value based on the energy balance of the propellant charge at the moment of projectile muzzle exit. Interior ballistics is the science that studies the processes leading to the movement of a projectile inside the weapon barrel under the action of propellant gases formed by the combustion of the propellant charge. The totality of phenomena occurring in the barrel from the moment of propellant charge ignition to the moment the outflow of propellant gases from the barrel ends after projectile exit is called the firing process. High pressure and temperature values of the propellant gases, along with a short duration, characterize the firing process. The firing process is divided into four periods: the preparatory (pyrostatic) period, the first (pyrodynamic) period, the second (thermodynamic) period, and the period of muzzle after-effect period [1].

The period of time from the exit of the projectile until the moment when the projectile leaves the stream of powder gases is called the period of muzzle after-effect period of the propellant gases on the projectile, while the period up to the moment when when the pressure of the powder gases equalizes with the atmospheric pressure is called the period of muzzle after-effect period of the propellant gases.

Corresponding author: S. Orašanin;
E-mail: orsasniper@gmail.com
DOI: 10.5937/OTEH26045O

Developing calculation methods for the projectile's motion parameters is the primary task of interior ballistics. When solving the primary task of interior ballistics, the first three periods are considered, while the period muzzle after-effect period of the propellant gases is not considered [2]. In classical methods of interior ballistics (IB), introducing the concept of the so-called mean ballistic pressure (MBP) into calculations is customary, whereby it is further assumed that both the pressure and temperature of the propellant gases at a given moment in time are equal throughout the entire working volume of the barrel bore.The concept of mean ballistic pressure is defined such that its work is equal to the total work of the propellant gases in the barrel, and that at a specific moment, it is equal at all points of the space behind the projectile. This pressure is called the mean ballistic pressure (MBP), assuming that the propellant burns under this mean pressure [1].The focus of this research is to determine the applicability of simplified analytical models for calculating mean ballistic pressure (MBP) using conventional interior ballistics (IB) formulas for engineering purposes. Their verification will be carried out by comparing the calculated MBP results with the experimental mean values of pressures measured by piezoelectric pressure sensors. The calculated MBP result was obtained using simplified analytical IB models based on the energy balance of the propellant charge for the case of complete combustion of the propellant charge at the moment of projectile muzzle exit.

2. ANALYTICAL MODELS

Two analytical models of classical interior ballistics will be briefly described, based on: Resal's equation (for calculating the mean ballistic pressure) and the formulas of Piobert and Mamontov (for calculating the pressure distribution along the barrel) at the moment of projectile muzzle exit, assuming complete combustion of the propellant charge.

The main assumptions under which these analytical models were derived are that the firing process is:

-adiabatic

-that the pressure, temperature, and density of the propellant gas mixture have identical values across all parts of the volume behind the projectile base at any instant of time t (the quasi-steadiness / quasi-stationary hypothesis).

-the composition of the combustion products does not change during firing, and the specific heat capacity of the gases at constant pressure $c_p=k\cdot R/(k-1)$ and constant volume $c_w=R/(k-1)$ remains the same and has an average value for the entire range of temperature variation; consequently, the adiabatic index $k=c_p/c_w$=const. Also, the gas constant (R) is invariant (constant) throughout the process

-The velocity of the propellant gases in the volume behind the projectile along the barrel varies linearly from zero at the breech to the velocity of the projectile, and the gas flow is one-dimensional.

-The barrel is cylindrical; the expansion of the powder chamber is neglected.

-The barrel remains stationary during firing.

2.1. Resal's Equation – Energy Balance Equation

The calculation of the mean ballistic pressure in the barrel at the moment of muzzle exit is based on the fundamental equation of interior ballistics. The French engineer Henri Resal first derived this equation in 1864. He based it on the first law of thermodynamics regarding the conservation of energy. This law applies to a thermodynamic system that includes both mechanical and thermal energy. The relation was named **Resal's equation** after its author.

Resal's equation applies to:

-Closed thermodynamic systems (energy exchange with the surroundings)

-Open thermodynamic systems (energy and mass exchange with the surroundings) [3].

In internal ballistics, Resal's equation represents the fundamental energy balance equation (the law of conservation of energy). It links the released energy of the propellant combustion and the potential energy of the propellant gases with mechanical work-the main and useful part of which equals the kinetic energy of the projectile-and losses within the firearm barrel.The mathematical formulation of Resal's equation, applied to the case of complete combustion of the propellant charge in the barrel at the moment of projectile muzzle exit before the onset of gas efflux, is:

$$Q= U +\varphi\cdot E_o \quad (1)$$

The total amount of heat (Q) released by the complete combustion of the propellant charge in the firing process

$$Q= \omega \times Q_{sp} \text{ [J]} \quad (2)$$

Unused internal energy of the propellant gases:

$$U =pW/(k-1) \text{ [J]} \quad (3)$$

The basic or useful mechanical work is the energy imparted to the projectile for its motion and is expressed through the kinetic energy of the projectile (muzzle kinetic energy of the projectile, E_0) plus the total secondary work and losses (φ).

$$\varphi \cdot E_o = \varphi \left(\frac{1}{2} q\cdot V_o^2\right) \text{ [J]} \quad (4)$$

The symbols used in the equations are:

ω [kg] - mass of propellant charge

Q_{SP} [J/kg] - specific energy of the propellant

q [kg] - projectile mass

V_0 [m/s] -projectile muzzle velocity

$\varphi=K + \frac{1}{3}\omega/q$ - coefficient of projectile fictitious mass which takes into account the total secondary work of the propellant gases in the barrel (work expended on the rotational motion of the projectile, overcoming friction, movement of gas);

K = 1,1 - coefficient which, according to Slukhotsky, has this value for small arms [1]

W[m^3]- total working (free) volume of the bore behind the projectile at the moment it leaves the muzzle

k- adiabatic index; for small arms propellants, a value of k = 1.25 is used [1]

The value of the mean ballistic pressure in the barrel volume at the initial moment of outflow can be analytically approximated, for the case of complete combustion of the propellant charge in the barrel bore, from expression (3) for the unutilized internal energy

$$p=(k-1)\cdot U/W \quad (5)$$

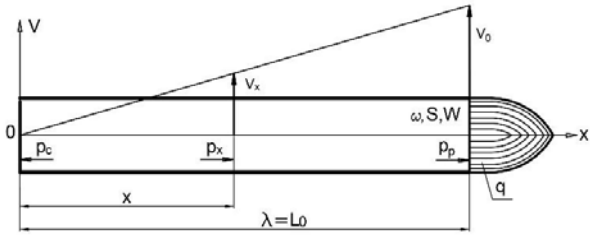


Figure 1. Schematic representation of the barrel at the moment the projectile leaves the muzzle

2.2.Piobert's Formula – Propellant Gas Pressure Distribution Along the Gun Bore Axis

During projectile motion, the propellant gases have different velocities at various cross-sections within the gun bore. These velocities range from zero at the breech end of the bore to the projectile velocity in the gas layer adjacent to the projectile base. Due to the accelerated gas flow through the bore volume, the pressure varies across different cross-sections of the barrel. The pressure distribution in the sections behind the projectile base depends on multiple factors. A functional relationship incorporating all of them would be highly complex and impractical for engineering calculations

Therefore, in conventional interior ballistics, the pressure distribution of propellant gases along the bore axis between the stationary breech face and the moving projectile is calculated using simplified equations, often referred to as Piobert's formulas. These formulas establish a relationship between the mean gas pressure (p), the breech pressure (p_C), and the projectile base pressure (p_p). The relationship between the mean ballistic pressure (p) and the projectile base pressure (p_p) is defined by the following expression:

$$p=p_p\left(1+\frac{1}{3}\frac{\omega}{\varphi_1 q}\right) \quad (6)$$

φ_1=1.05 - The value of the partial coefficient of fictivity for small arms ammunition with a full metal jacket (FMJ) lead-core projectile [1].

Equation (6) is Piobert's formula (French general and scientist Guillaume Piobert), which accounts for the forces resisting the projectile's motion (φ_1), as proposed by the Russian engineer P. N. Shkvornikov [1].
The relationship between the pressure at the breech face (p_c) and the pressure at the base of the projectile (p_p) is determined by the expression:

$$p_c=p_p\cdot\left(1+\frac{1}{2}\frac{\omega}{\varphi_1 q}\right) \quad (7)$$

A contribution to the further upgrade of Piobert's formula was made by the Russian engineer M. A. Mamontov [4]. He correlated the dependence between the pressure at a distance "x" from the breech face and the pressure at the base of the projectile (p_p) (Figure 1). The formula was applied to the case when the combustion of the propellant charge is completed at the moment the projectile leaves the muzzle, and is given by the expression:

$$p_x=\left[1+\frac{1}{2}\frac{\omega}{\varphi_1 q}(1-z^2)\right]p_p \qquad z=x/\lambda \quad (8)$$

3. RESULTS

In order to compare the calculated results with experimental data, bore pressure measurements of the propellant gases must be conducted. An EPVAT*(Electronic Pressure Velocity and Action Time)* 7.62x51mm test barrel was selected as the test case. Firing was conducted using 7.62x51mm M80 FMJ BT 9.4g/145gr reference ammunition, manufactured by Prvi Partizan AD, Užice. The ammunition was loaded with propellant manufactured by Milan Blagojević-Namenska AD, Lučani. Ammunition and propellant data were provided by the manufacturers. Two measurement ports for piezoelectric transducers were configured on the 7.62x51mm EPVAT test barrel in accordance with NATO STANDARD AEP-97 [5] to record pressure-time curves. The measurement port on the test barrel (Figure 2.) with abscissa x_1, as specified by standard, records the maximum barrel pressure (p_m), while the measurement port with abscissa x_2 records the maximum gas port pressure (p_g).Firing was conducted from the test stand using the selected test barrel.

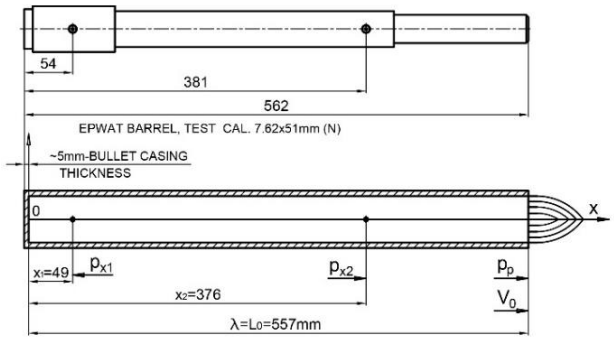


Figure 2. Location of piezoelectric transducers on the EPVAT test barrel (top) and schematic diagram of the barrel with abscissas of the gauge positions in the barrel bore (bottom)

The measurement system is initiated upon firing, precisely at the moment the firing pin strikes the cartridge primer. For each round fired from the selected test barrel, pressure was measured at two gauge ports, and the projectile muzzle velocity was recorded. The projectile exit time from the barrel (t_{ex}) was measured using a stop probe, and the corresponding pressure value at that specific instant was read from the recorded pressure curve. Pressure measurements were performed using a Kistler 6215 piezoelectric transducer. The muzzle velocity

was measured at a distance of 2.5 meters from the muzzle. The measurements and test-firings were conducted in the ballistic tunnel of the Agency for Testing, Stamping and Marking of Weapons, Devices and Ammunition in Kragujevac.

Table 1 lists the data for the selected cartridge (q, ω , V_0), propellant (Q_{sp}, k), data for the selected test barrel-specifically the barrel bore volume (W)-and the value of the total coefficient of projectile fictitious mass (φ)for small arms, as defined by classical interior ballistics (IB). The barrel bore volume (W) comprises the volume of the main bore (bore guide) and the internal volume of the empty cartridge case inserted into the barrel chamber.

The specified data are sufficient to calculate the mean ballistic pressure (MBT) in the barrel at the moment the projectile leaves the muzzle. By substituting the stated data into expressions (1)–(5), the theoretical value of the MBP was calculated in Table 1, amounting to 616 x10^5 [Pa].Table 2. presents the mean values for twenty indipedent firing measurements: the peak pressures for both gauge locations and the measured projectile velocity values extrapolated to the muzzle. The measured projectile velocities and peak pressures at both gauge locations met the criteria regarding both the magnitude of the measured values and the dispersion of the results; therefore, the ammunition lot can be considered a representative sample. The composite diagram of the p(t) pressure curves for both gauge locations is shown in Figure 3.

Table 1. Mean barrel pressure (p) values based on the energy balance for the 7.62x51mm M80 FMJ BT 145 gr PPU cartridge.

Projectile mass q ×10^{-3} [kg]	9.4
Mass of powder charge: ω×10^{-3} [kg]	2.94
Length of test tube L_c [m]	0.56
Bore cross-sectional area S×10^{-4} [m^2]	0.4571
Total working (free) volume of the bore W×10^{-4} [m^3]	28
Muzzle velocity V_0 [m / s]	840
Muzzle kinetic energy of the projectile E_0 =(q·V^2)/2 [J]	3316
Propellant: Spherical, Milan Blagojević-Lučani	SB520
Specific energy of the propellant Q_{sp} [J / g]	3700
Total energy of propellent charge: ω·Qsp [J]	10878
Total coefficient of coefficient of projectile fictitious mass: φ= K + $\frac{1}{3}$ ω/q; K=1.1	φ =1.2
Fictitious muzzle energy of the projectile: φ·E_o [J]	3979
Internal energy of propellant gases in the barrel bore: U= ω·Q_{sp} - φ·E_o [J]	6899
Adiabatic index; for small arms propellants: k	1.25
Calculated mean ballistic pressure in the bore at muzzle exit: p_o=(k-1)·U/W ×10^5 [P a]	**616**

In order to compare the obtained pressure measurement results (Table 2) with the theoretical MBP (Mean Ballistic Pressure) value calculated via the propellant charge energy balance, it is necessary to recalculate the measured values using the Piobert and Mamontov formulas. In formula (8), pressure p_x is a known, measured value, while the pressure at the projectile base (p_p), i.e., at the muzzle, as well as the mean ballistic pressure (p), must be determined from formulas (8) and (6). Therefore, by correcting the measured pressure values, the pressure values at the projectile base and the mean ballistic pressure (p) were determined and are presented in Table 3 for both measurement locations.

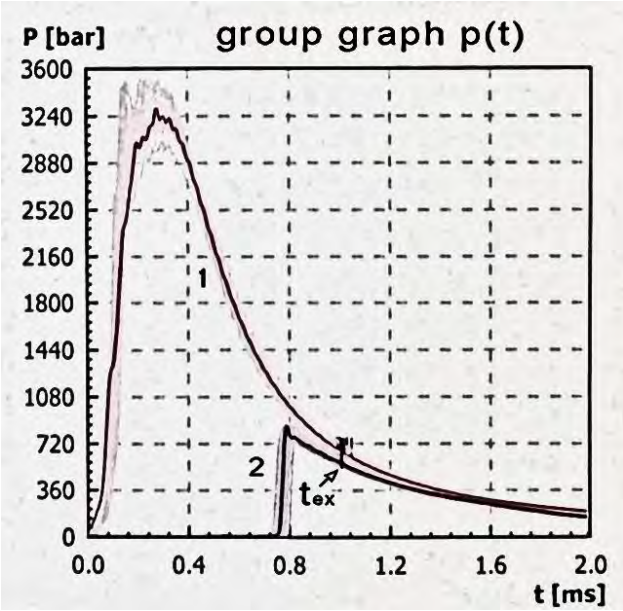


Figure 3. Combined plot of pressure-time curves p(t) for both gauge locations: x_1 (1) and x_2 (2). The projectile muzzle exit time (t_{ex}) was determined by a stop probe and is indicated on the graph by a vertical tick mark. The bold curves represent the mean pressure curves from twenty measurements.

Table 2. Mean measured pressure values at measuring points x_1 and x_2

Measuring point **(Figure 2.)** x [mm]	x_1=49	x_2=376
Mean values of measured max pressures p_m ×10^5[Pa]	p_{m1}=3298.6	p_{m2}=848.6
Average measured velocity $V_{2.5}$ [m / s]	838	
Muzzle velocity V_0 [m / s]	840	
Pressure curve reading at muzzle exit p ×10^5[Pa]	P_{x1}=656.4	P_{x2}=564.8
Test barrel bore length L_0= λ [mm]	557	557
Relative coordinate z=x/λ	z_1=0.08797;	z_2=0.67504

Table 3. Projectile base pressure (p_p) and mean ballistic pressure (p) values obtained by recalculating measurement results at gauge locations $x_1(p_m)$ and $x_2(p_p)$.

Pressure gauges positions in the barrel bore x [mm]	x_1=49	x_2=376
Test barrel bore length $L_0 = \lambda$ [mm]	557	557
Pressure curve reading at muzzle exit $p_x \cdot 10^5$[Pa]	**p_{x1}=656.4**	**p_{x2}=564.8**
Relative coordinate $z=x/\lambda$; z^2 $1-z^2$	z_1=0.08797; z_1^2=0.007739 $1-z^2$=0.9923	z_2=0.67504 z_2^2=0.4557 $1-z_2^2$=0.5443
$\frac{1}{2}\frac{\omega}{\varphi_1 q}$ φ_1 =1.05 $\omega=2.94\times10^{-3}$ [kg]; $q=9.4\times10^{-3}$[kg]	0.1564	0.1564
$p_x=[1+\frac{1}{2}\frac{\omega}{\varphi_1 q}(1-z_2^2)]p_p$ Formula (8)	p_{x1} $=1.1552\cdot p_{p1}$	p_{x2} $=1.08513\cdot p_{p2}$
$p_p= p_x/[1+\frac{1}{2}\frac{\omega}{\varphi_1 q}(1-z_2^2)]$	p_{p1} $=p_{x1}/1.1552$	p_{p2} $=p_{x2}/1.08513$
The projectile base pressure: $p_p \cdot 10^5$[Pa]	**p_{p1}=568.2**	**p_{p2}=520.5**
$\frac{1}{3}\frac{\omega}{\varphi_1 q}$	0.09929	0.09929
$p=\left(1+\frac{1}{3}\frac{\omega}{\varphi_1 q}\right)\cdot p_p$ Formula (6)	p_1= 1.09929$\cdot p_{p1}$	p_2= 1.09929$\cdot p_{p2}$
Calculated mean ballistic pressure in the bore at muzzle exit $p \cdot 10^5$[Pa]	**p_1=624.6**	**p_2=572**

4. DISCUSSION

The value of the mean ballistic pressure (MBP) calculated (Table 1) via the propellant charge energy balance (616 × 10^5 [Pa]) practically matches the MBP value obtained by correcting the mean measured value at gauge station x_1 (624.6 × 10^5 [Pa]), being lower by 1.4% (Table 4). When compared to the mean pressure obtained by correcting the mean measured value at gauge station x_2 (572 × 10^5 [Pa]), it is higher by 7.7% (Table 4). The mean ballistic pressure (MBP) calculated using the analytical propellant charge energy-balance method is in very good agreement with the reference value at the first gauge station, (x_1), where the pressure difference is negligible, while at the second gauge station, (x_2), the difference remains within an acceptable range.

The advantage of the analytical method for calculating the mean ballistic pressure (MBP) is its simplicity, as it does not require a detailed knowledge of propellant properties and the design characteristics of the projectile, weapon, and barrel, which are otherwise essential for a reliable internal ballistics (IB) calculation (Table 1).

All the previously mentioned simplifications and assumptions introduced herein aim to streamline the mathematical model and make it suitable for practical engineering calculations

Table 4. Comparative values of theoretical and empirical mean ballistic pressures (MBP) in the gun barrel bore

	Measuring point (**Figure 2.**)	x_1	x_2
4.1.	Pressure curve reading at muzzle exit $p_x \cdot 10^5$[P a]	656.4	564.8
4.2.	Mean ballistic pressure obtained by correcting the read value. $p_x \cdot 10^5$[P a]	**624.6**	**572**
4.3.	Calculated mean ballistic pressure in the bore at muzzle exit $p_0 \cdot 10^5$[P a]	**616**	**616**
4.4.	The relative difference between MBP (4.2) calculated in the bore at muzzle exit (p_0) and (4.3) obtained by correcting the read value (p_x)	-1.4%	+7.7%

In classical interior ballistics methods, it is common practice to neglect phenomena that do not significantly affect calculation accuracy and whose detailed consideration would greatly complicate the computation. Instead, these phenomena are accounted for using simple empirical formulas derived from experimental test results of similar weapon and ammunition models.

5. CONCLUSION

This paper addresses the evaluation of the mean ballistic pressure (MBP) in the bore behind the projectile of a small arms weapon at the instant of muzzle exit, utilizing the energy balance of the propellant charge under the condition of completed propellant combustion. The following conclusions can be formulated:

-Analytical expressions derived within classical ballistics, incorporating specific assumptions and relevant simplifications, enable the computation of the MBP based on the propellant charge energy balance and the pressure variation in the volume between the breech and the moving projectile;

-The derived theoretical MBP value, compared against empirical data measured via piezoelectric transducers, proves acceptable and sufficiently precise for engineering applications;

-The theoretical MBP value can serve as an initial/input value for the engineering design and calculation of muzzle devices (flash suppressors, muzzle brakes, recoil boosters, silencers);

-The primary advantage of the demonstrated analytical analytical method for MBP calculation is its simplicity, as it bypasses the need for in-depth data on propellant properties and the structural parameters of the

weapon, barrel, and ammunition required for reliable internal ballistics modeling.

References

[1] M. E СЕРЕБРЯКОВ, Внутреняя баллистика ствольных систем и пороховых ракет, ОБОРОНГИЗ Москва, 1951

[2] E. B. ЧУРБАНОВ: Внутреняя баллистика, учебник, Военная Артиллерийская,, Ордена Ленина Краснознаменная Академия, им. М.И. Калинина, Ленинград, 1975

[3] M.C. ГОРОХОВ: Внутреняя баллистика ствольных систем, Москва, 1985

[4] M.A. MAMOHTOB: Некаторые случаи течения гаса по трубам, насадкам и проточним сосудам, Г.И.О.П., Москва, 1951

[5] NATO STANDARD AEP-97: MULTI-CALIBRE MANUAL OF PROOF AND INSPECTION (M-CMOPI) FOR NATO SMALL ARMS AMMUNITION, NSO 2020

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DEFINITION OF A METHOD FOR VERIFYING THE QUALITATIVE CHARACTERISTICS OF ELECTRIC IGNITER HEADS

JELENA PETKOVIĆ-CVETKOVIĆ
Military Technical Institute, Belgrade, jpetkovic82@gmail.com

BRANKA DŽUNIĆ
Military Technical Institute, Belgrade, brankadz99@gmail.com

SLOBODAN BOSILJČIĆ
Military Technical Institute, Belgrade, bobanosiljcic@yahoo.com

GORAN MARJANOVIĆ
Military Technical Institute, Belgrade, gmarjanovic1961@gmail.com

DRAGAN KNEŽEVIĆ
Military Technical Institute, Belgrade, dragankn@gmail.com,

JELENA MOJSILOVIĆ
Military Technical Institute, Belgrade, mojsilovi.jeca@gmail.com

MARKO RADISAVLJEVIĆ
Military Technical Institute, Belgrade, markoradisavljevic.94@gmail.com

Abstract: *Most pyrotechnic chains begin with electrical ignition using an electric igniter head. The reliability of electrical igniter heads (EIHs) represents one of the key factors governing the proper functioning of a pyrotechnic chain, since the mere occurrence of successful initiation is not sufficient for a complete assessment of the quality and functional compatibility of the igniter heads. The aim of this study was to develop an experimental method for the quantitative characterization and comparative evaluation of EIHs based on their pressure–time response. The tests were performed in a manometric chamber under controlled conditions, and the qualitative characteristics were determined using the maximum pressure, Pmax, ignition delay time, Tkp, time to maximum pressure, Tmax, and indicators of result variability within a tested series. Seven different types of EIHs were investigated. The obtained results enabled their comparative classification with respect to maximum pressure intensity, temporal characteristics, and repeatability. The study demonstrates that the combined characterization of maximum pressure, temporal response, and statistical variability provides a suitable basis for a more objective assessment of the quality and functional compatibility of electrical igniter heads.*

Keywords: *Electric igniter head, Manometric chamber, Maximum pressure, Delay time, Time to reach maximum pressure.*

1. INTRODUCTION

Electrical pyrotechnic initiators represent the initial elements of numerous pyrotechnic systems, and their primary function is to convert electrical energy into the thermal and chemical energy required to initiate a pyrotechnic chain. Their application encompasses a wide range of systems in which reliable and temporally precise initiation is required. Consequently, electrical initiators are expected to exhibit high reliability, reproducibility, and stable performance throughout their intended service life [1,2].

The mere fact that initiation was successfully accomplished is not always sufficient for a full assessment of the quality and functional compatibility of the initiator. Of particular importance are those characteristics that determine the response time of the initiator and the intensity of energy release, as deviations in these characteristics may affect the operation of subsequent elements in the pyrotechnic chain.

During an adaptation of a fuze for a new warhead model, certain deficiencies were observed with the use of the standardized electric igniter. During a series of tests, the EIH occasionally exhibited an inadequate response. It was assumed that an excessive amount of combustion products and soot could cause mechanical damage to the fuze components, resulting in improper and unreliable operation. Therefore, it was necessary to identify a

Corresponding author: J.Petković Cvetković
E-mail: jpetkovic82@gmail.com

DOI: 10.5937/OTEH26046P

suitable EIH capable of providing a more uniform energy release and a lower amount of soot in the combustion products. Since functional tests performed on the complete adapted fuze and tandem shaped-charge warhead assembly are complex and involve numerous variables, it was necessary to establish a method that would enable the rapid and reliable characterization of electrical igniter heads. [3,4]

The operation of an electrical initiator is based on the heating of an electrical initiating element, i.e., a bridgewire, by the passage of electric current. The released heat is transferred to the primary pyrotechnic charge and, once the temperature required for initiation is reached, an exothermic chemical reaction is initiated. This reaction results in rapid energy release and the formation of gaseous and condensed reaction products. [5]

When an initiator is tested in a closed volume, the gaseous reaction products cause an increase in pressure that can be recorded as a function of time. In this manner, the complex initiation process can be characterized on the basis of a measurable pressure–time response. Jang et al. [6] reported that the time interval between the application of electrical current and the onset of pressure generation can be defined as the ignition delay time, whereas the maximum pressure and the time required to reaching this pressure are associated with the combustion characteristics of the pyrotechnic material. [7]

The pressure–time relationship represents an important source of information regarding initiator behavior. Maximum pressure represents a measure of the intensity of the pressure response, whereas the time required to reach the maximum pressure provides information regarding the dynamics of the process. The ignition delay time, on the other hand, characterizes the interval between the application of the electrical stimulus and the onset of pressure development. According to the results reported by Jang et al., the ignition delay is primarily associated with the condition of the bridgewire and the contact between the bridgewire and the primary pyrotechnic charge, whereas the maximum pressure and the time required to reach it are more closely related to the combustion characteristics of the pyrotechnic material. [8]

The performance of an initiator is influenced not only by its design and the composition of the pyrotechnic charge but also by the conditions to which the initiator is exposed before and during testing. Temperature is one of the important factors that may affect the reaction rate and pressure development. [8]

In addition to the mean values of the characteristic parameters, reproducibility of the results is of great importance for quality assessment. Under practical manufacturing conditions, a certain degree of variability is unavoidable, meaning that individual initiators of the same type may not produce completely identical responses. Therefore, in addition to determining the mean values of the measured parameters, their dispersion should also be evaluated. Standard deviation is one of the basic measures of variability, while the coefficient of variation can be used to compare the relative dispersion of results between different groups.

Considering the above, reliable verification of the qualitative characteristics of electric igniter heads requires a method that simultaneously addresses time response, reaction intensity, pressure-development dynamics, and reproducibility of the results. Accordingly, the principal parameters considered for characterization are ignition delay time (t_{kp}), maximum pressure (P_{max}), maximum pressure arrival time (t_{max}), and the standard deviation of the results within the tested series.

The aim of this study is to define a method for verifying the qualitative characteristics of electric igniter heads based on their pressure-time response recorded in a pressure measurement chamber. Five different types of electric igniter heads will be investigated under controlled testing conditions. The obtained results will be used to analyze the selected characteristic parameters, compare their values, and evaluate the dispersion of results within each series. In this way, the study seeks to establish a method that enables an objective assessment of the quality and reproducibility of electric igniter heads and provides an appropriate basis for their verification.

2. MATERIALS AND METHODS

2.1. Materials

The experimental investigation was conducted using different types of electrical igniter heads (EIHs) in order to determine and compare their combustion characteristics. Five different types of electric igniter heads were investigated under controlled testing conditions and at a loading density of 0.01 g/cm³. In the initial phase of the investigation, standard electrical igniter labeled as MB-2N, MB-2NP1, and DISV types were analyzed. In the second phase, two series of modified EIH, MB-2N M and DISV M, were tested.

The investigated EIH types were selected because of differences in their design and manufacturing characteristics, which may result in differences in the intensity and dynamics of pressure development. The MB-2N and MB-2NP1 igniter heads are manufactured individually using a predominantly manual process, whereas DISV igniter heads are produced in comb-like assemblies, allowing entire series to be simultaneously immersed in a slurry of the initiating pyrotechnic composition. Additionally, these igniter heads feature different initial pyrotechnic mixture compositions. The MB-2N and MB-2NP1 igniter heads are based on potassium perchlorate and Lead (II) thiocyanate, and DISV based on Zirconium and Lead (II) oxide.

2.2. Testing Method

The electrical circuit resistance was measured for all samples from the five investigated EIH series with OM200 M, Trio Electronics Ltd, and EIHs with closely comparable resistance values were selected within each series in order to minimize the influence of electrical resistance on the ignition rate and, consequently, on the ignition delay time.

The characteristics of the EIHs were determined using a

closed manometric bomb, based on the measurement of pressure changes within a confined volume following activation of the electrical igniter head. The products generated during the pyrotechnic reaction produce an increase in pressure inside the chamber, and this pressure change is recorded as a function of time.

The application of the manometric method enables the behaviour of an EIH to be characterized by means of its pressure-time curve. This approach allows simultaneous analysis of the intensity of the generated pressure, the time response of the igniter, and the rate of pressure development. A similar approach has been applied in previous studies of electric pyrotechnic initiators, in which the Closed Bomb test was used to characterize initiator performance on the basis of the recorded pressure-time response. [9]

The pressure transducers used in the present study were designed for the measurement of static and dynamic pressures in liquids and gases. They are available in various measuring ranges, from 1 to 3000 bar. Transducers with measuring ranges between 0 and 3000 bar operate according to the strain-gauge principle. The strain gauge is located on the side of the measuring tube opposite the measurement medium and is enclosed within a hermetically sealed reference chamber, thereby protecting it from external environmental influences.

All five EIH series were tested in a modified manometric bomb [10] Figure 1., with a free volume of 2 cm^3, corresponding to a loading density of approximately 0.01g/cm^3. The setup was equipped with a calibrated PZ|IC pressure transducer with a measuring range of up to 200 bar.

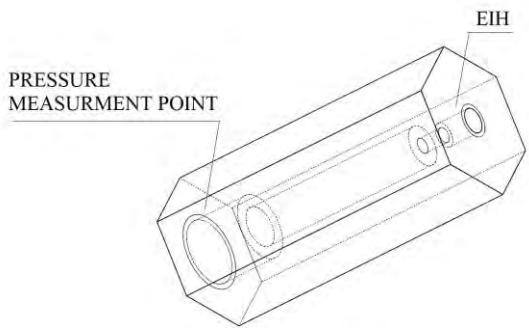


Figure 1. Modified manometric bomb used for testing electrical igniter heads.

The transducer housing was made of stainless steel and hermetically sealed the measuring system, thereby protecting it from harmful external influences and ensuring that the reliability and accuracy of the transducer were not impaired even under demanding operating conditions.

During the first phase of the EIH investigation, three types of igniter heads were assembled into the holders of the modified manometric bomb. The first series consisted of standardized MB-2N EIHs, for which deficiencies had previously been observed during operation in an adapted warhead fuze. The second series consisted of non-standardized DISV EIHs. Owing to their low production cost and high reliability, these EIHs are used in initiation chains during internal investigations of various pyrotechnic elements. The third series consisted of MB-2NP1 EIHs, which are also standardized EIHs. They are manufactured using the same technology as MB-2N EIHs, but differ in the composition of the initiating pyrotechnic mixture.

The investigated EIHs were subsequently modified. First, the MB-2N EIH was modified, designated MB-2N M, by reducing the amount of initiating pyrotechnic composition applied to the bridgewire. Specifically, the igniter head was immersed in the initiating composition once instead of twice.

The DISV EIH was modified, designated DISV M, only in terms of its design, without changing the composition of the initiating pyrotechnic mixture. A protective sleeve was designed and placed around the igniter head in order to provide a directed effect required for its application in warhead fuzes.

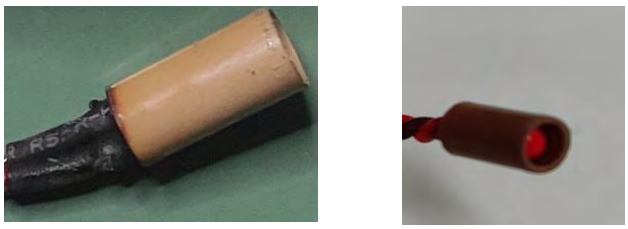


Figure 2. Modified, DISV-M and MB-2N M, electrical igniter heads.

2.3. Determination of Characteristic Parameters

The pressure-time response obtained during testing of the electrical igniter heads was characterized by four time parameters: T_{start}, T_{max}, T_{km}, and T_{kp}. T_{start} denotes the onset of the recorded pressure response, whereas T_{max} corresponds to the time at which the maximum pressure, P_{max}, is reached. The pressure-development time, T_{km}, was calculated as:

$$T_{km}=T_{max} - T_{start}$$

Unlike T_{start} and T_{max}, which represent absolute time coordinates, T_{km} represents the duration of the pressure-development phase from the onset of the pressure response to the attainment of P_{max}.

T_{kp} represents the ignition delay time, defined as the time interval between application of the electrical activation pulse and the onset of the recorded pressure response. If (T_0) denotes the reference time corresponding to the application of the electrical pulse, T_{kp} is calculated as:

$$T_{kp}=T_{start} - T_0$$

where T_0 denotes the reference time corresponding to the application of the electrical pulse.

For the investigated EIH types, an extensive set of statistical methods is not required. The basic statistical analysis includes determination of the mean value, standard deviation, and coefficient of variation.
Mean value:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Standard deviation:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Coefficient of variation:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

3. RESULTS AND DISCUSION

For all samples belonging to the five investigated EIH series, the electrical circuit resistance was measured, and EIHs within a narrow resistance range were selected in order to minimize its influence on the ignition rate and, consequently, on the ignition delay time. Table 1 presents the applicable resistance requirements according to the relevant internal standards, together with the measured resistance ranges of the tested samples.

Table 1. Required and measured resistance range

EIH type	Required resistance range [Ω]	Measured resistance range [Ω]
MB-2N	2,5-4,5	2,92-3,31
MB-2NP1	0,6-1,2	0,93-1,06
DISV	1,5-2,5*	1,75 – 1,94
MB-2N M	2,5-4,5	2,92-3,31
DISV M	1,5-2,5*	1,67 – 1,91

*For DISV EIHs, the resistance value is not defined by a formal standard; the indicated range represents the range used for sample selection during the present investigation.

Tables 2–4 present the measured values of maximum combustion pressure, ignition delay time, and pressure-development time. For the first three EIH groups, MB-2N, MB-2NP1, and DISV. The pressure variation as a function of time is also presented graphically in Figures 3 for the MB-2N EIHs.

Table 2. Maximum combustion pressure, ignition delay time, and pressure-development time for the MB-2N EIHs

	T_{kp}[s]	T_{km}[s]	P_{max}[bar]
MB2N_1	0,0054	0,0099	10,8780
MB2N_2	0,0053	0,0095	10,9207
MB2N_3	0,0108	0,0099	9,5658
MB2N_4	0,0069	0,0104	11,7066
MB2N_5	0,0081	0,0103	14,8482
MB2N_6	0,0060	0,0102	12,9888
MB2N_7	0,0066	0,0100	10,0929
Mean value	0,0070	0,0100	11,5716
St.deviation	0,0019	0,0003	3,3129

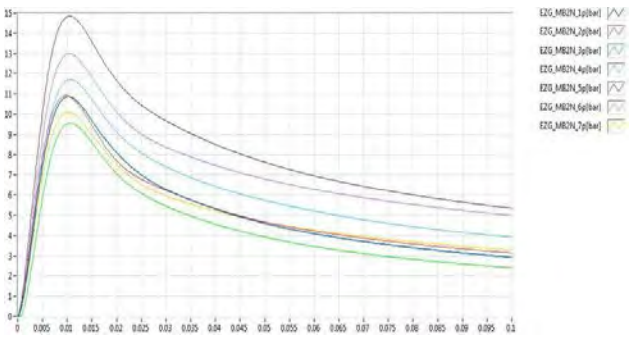


Figure 3. The pressure variation as a function of time for the MB-2N EIHs

The measured maximum pressure values for the MB-2N EIHs ranged from 9,56 to 14,85 bar, with a mean value of 11,57 bar and a standard deviation of 3,31. The range of the results indicates relatively pronounced variability in the maximum pressure among individual samples. A coefficient of variation of approximately 28.6% further demonstrates that the differences between the samples are not negligible relative to the mean response level obtained. This variability is likely attributable to the manufacturing process, specifically to variations in the amount of pyrotechnic composition applied to the initiator element, which most likely points to the underlying causes of failures and problems encountered during its application in the fuze of the warhead.

From the perspective of quality control, this result demonstrates that the behavior of the tested samples cannot be characterized solely by the mean value; the stability of the response from sample to sample must also be considered. In contrast to the observed variation in Pmax, the pressure-development times were highly consistent, ranging from 0,95ms to 1,04 ms with a standard deviation of 0,0003. This is the result of a narrow resistance range of selected group.

Table 3. Maximum combustion pressure, ignition delay time, and pressure-development time for the MB-2NP1 EIHs

EIH type	T_{kp} [s]	T_{km} [s]	P_{max}[bar]
MB2NP1_1	0,0016	0,0016	9,055
MB2NP1_2	0,0005	0,0015	11,675
MB2NP1_3	0,0007	0,0045	7,711
MB2NP1_4	0,0014	0,0013	8,952
MB2NP1_5	0,0026	0,0052	7,212
Mean value	0,0014	0,00282	8,920
St.deviation	0,00018	0,00187	1,728

The maximum pressure values obtained for the MB-2NP1 EIHs ranged from 7.21 to 11.68 bar, with a mean value of 8.92 bar and a standard deviation of 1.73. The corresponding coefficient of variation was approximately 19.4%, indicating that the differences among individual samples were not negligible relative to the mean response level.

Table 4. Maximum combustion pressure, ignition delay time, and pressure-development time for the DISV EIHs

EIH type	T_{kp} [s]	T_{km} [s]	P_{max}[bar]
DISV_1	0,0062	0,0230	4,4128
DISV_2	0,0114	0,0180	4,6951
DISV_3	0,0061	0,0280	3,3053
DISV_4	0,0054	0,0216	5,4752
DISV_5	0,0095	0,0241	2,3857
Mean value	0,0077	0,0230	4,0548
St.deviation	0,0026	0,0036	1,4759

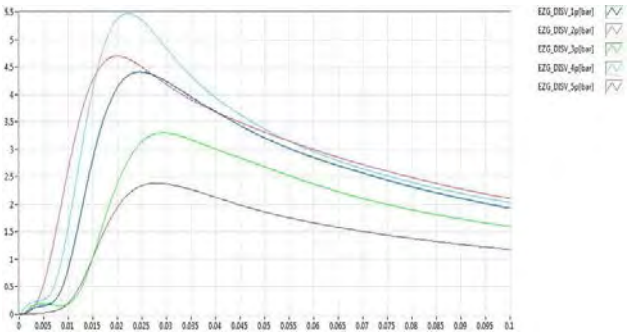


Figure 4. The pressure variation as a function of time for the DISV EIHs

The measured maximum combustion pressure values for the DISV EIHs ranged from 3,305 to 5,475 bar, with a mean value of 4,05 bar and a standard deviation of 1,47. In contrast to the observed variation in Pmax, the pressure-development times were highly consistent, ranging from 1,8 ms to 2,8 ms with a standard deviation of 0,0036. This is the result of a narrow resistance range of selected group.

Following these investigations, tests involving DISV EIHs installed in adapted tandem shaped-charge warhead fuzes were performed. Several EIHs were damaged during assembly into the fuze and subsequently failed during functional testing. In order to eliminate this problem while simultaneously enabling the directed action required for application in the warhead fuze, a protective sleeve was designed and manufactured around the igniter head, resulting in the modified DISV M configuration.

Table 5 presents the measured parameters for the modified MB-2N M electrical igniter heads. Modification of the MB-2N design resulted in a reduction of the mean maximum pressure and the amount of soot produced, as expected, since the igniter head was immersed in the initiating pyrotechnic mixture once instead of twice.
However, the dispersion of the individual maximum pressure values remained relatively high, with a standard deviation of 1.70 and a coefficient of variation of approximately 20.42 %.

Table 5. Maximum combustion pressure, ignition delay time, and pressure-development time for the MB-2N M EIHs

EIH type	T_{kp} [s]	T_{km} [s]	P_{max}[bar]
MB2N M_1	0,0075	0,0039	10,95
MB2N M_2	0,0086	0,0024	9,17
MB2N M_3	0,0082	0,0017	7,56
MB2N M_4	0,0054	0,0026	7,91
MB2N M_5	0,006	0,0016	9,78
MB2N M_6	0,0108	0,0085	6,45
MB2N M_7	0,0083	0,0136	6,49
Mean value	0,0078	0,0049	8,33
St.deviation	0,001786	0,001786	1,70

Table 6. Maximum combustion pressure, ignition delay time, and pressure-development time for the DISV M EIHs

EIH type	T_{kp} [s]	T_{km} [s]	P_{max}[bar]
DISV-M_1	0,0104	0,0285	3,9482
DISV-M_2	0,0124	0,0378	2,3896
DISV-M_3	0,0127	0,0288	4,7043
DISV-M_4	0,0098	0,0351	3,5008
DISV-M_5	0,0125	0,0368	2,7811
Mean value	0,0116	0,0334	3,4648
St.deviation	0,0014	0,0045	0,8487

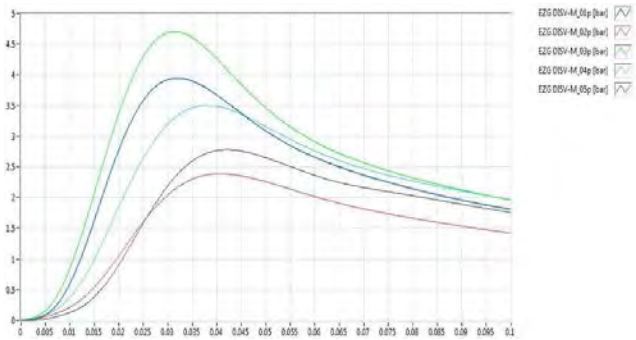


Figure 5. The pressure variation as a function of time for the DISV EIHs

The measured values for the DISV M EIH group were relatively uniform, ranging from 2,39 to 4,70 bar, with a mean value of 3,46, although the mean value was somewhat lower than that obtained for the original DISV configuration. This modification resulted in an EIH with a low soot content, low standard deviation 0,95, and simplified laboratory handling, making it suitable for application in the warhead fuze.

Since the modification and assembly of the DISV M EIH in the actual fuze configuration required an increase in the available free volume, EIHs with a higher energy output will be required in further investigations. Accordingly, future research should focus on modifying the composition of the initiating pyrotechnic mixture in order to achieve a higher energy output than that of the DISV EIH while producing fewer solid combustion residues than the modified MB-2N M configuration.

4. CONCLUSION

The study developed and experimentally verified a method for the multiparametric characterization of electrical igniter heads (EIHs), based on the analysis of pressure variation as a function of time under controlled manometric testing conditions. Testing of five EIH types, using the parameters P_{max}, t_{kp}, t_{max}, and result variability, confirmed the existence of measurable differences in response intensity, temporal dynamics, and repeatability. Thus, the combined analysis of pressure, temporal, and statistical parameters provides a more reliable basis for the assessment of EIH characteristics than the analysis of individual parameters alone. Integration testing within a real functional assembly confirmed that individual tests should be regarded as a necessary, but not sufficient, level of verification. The proposed method has the potential to provide more objective quality control, facilitate the comparison of design variants, and enable the early detection of deviations. Future research should focus on larger sample series, statistical validation of the proposed criteria, investigation of the effects of aging and environmental conditions, and the development of predictive models linking component-level parameters with system-level performance and reliability.

References

[1] CONKLING J. A., MOCELLA C. J.: Chemistry of Pyrotechnics, second edition, 2011.

[2] ORBOVIC N.: Uvod u energetske materije, KIZ „Centar“, 2020, Belgarde

[3] MATIJEVIĆ J.: Inicijalne pirotehničke smeše za zapaljive glavice, Magistarski rad, TMF, Univerzitet u Beogradu,1986.

[4] ĐURICA A.: Pirotehnika i inicijalna sredstva, Vojnoizdavački zavod, Beograd 1983.

[5] MATIJEVIĆ J.: Proučavanje inicijalnih pirotehničkih smeša za zapaljive glavice, int. dok., TI-950, Vojnotehnički institut, Beograd, 1988.

[6] JANG S.-G., et al.: Understanding aging characteristics of a pyrotechnic initiator through performance modeling in Closed Bomb Tests, Propellants, Explosives, Pyrotechnics, 49, 2024. DOI: 10.1002/prep.202300284.

[7] MATIJEVIĆ J.: Proučavanje inicijalnih pirotehničkih smeša za zapaljive glavice, int. dok., OP-266, Vojnotehnički institut, Beograd, 1990.

[8] VARGHESE P. L.: Investigation of energy transfer in the ignition mechanism of a NASA standard initiator,” NASA-CR-184673, 1988

[9] ULAS A., et al.: “An Investigation of the Performance of a Boron/Potassium-Nitrate Based Pyrotechnic Igniter,” Propellants, Explosives, Pyrotechnics, 31, 311–317, 2006. DOI: 10.1002/prep.200600047.].

[10] KRSTOVIĆ M., PETKOVIĆ-CVETKOVIĆ J., DIMITRIJEVIĆ I., ĆITIĆ A., BAJIĆ D., “Calorimetric bomb test as a method for the selection of optimal boron powder for pyrotechnic applications”, 27th Seminar on New Trends in Research of Energetic Materials - NTREM 2025

[11] MARJANOVIĆ G., BOSILJČIĆ S., PETKOVIĆ-CVETKOVIĆ J.: “ Ispitivanje elektronike TKBG I uticaja elektro zapaljivih glavicaMB-2N I DISV na funkciju adaptiranog upaljača” int. dok., VTI-03-01-0 568, Vojnotehnički institut, Beograd, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MODIFIED PBXN-5 EXPLOSIVE: EFFECT OF IF-WS_2 NANOPARTICLES ON PHLEGMATIZATION OF HMX

DANICA BAJIĆ
Military Technical Institute, Belgrade, Serbia, danica.bajic@mod.gov.rs

MIRJANA KRSTOVIĆ
Military Technical Institute, Belgrade, Serbia

IVAN DIMITRIJEVIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Serbia
MC Labor, Belgrade, Serbia

MLADEN TIMOTIJEVIĆ
Military Technical Institute, Belgrade, Serbia

OGNJEN RAJKOVIĆ
Military Technical Institute, Belgrade, Serbia

JELENA MARINKOVIĆ
Military Technical Institute, Belgrade, Serbia

Abstract: This study examines the influence of inorganic fullerene-like tungsten disulfide nanoparticles (IF-WS_2) on the rheological behavior of Viton A solutions and their application in the phlegmatization of octogen (HMX). IF-WS_2 nanoparticles were selected due to their unique quasi-spherical morphology, lubricating properties, thermal and mechanical resistance. Their incorporation into fluoropolymer binders like Viton A is expected to reduce interparticle friction and enhance energy dissipation under mechanical stimuli, thereby decreasing sensitivity to impact and friction. At the same time, IF-WS_2 nanoparticles influence the processing parameters such as viscosity and particle coating efficiency. Two polymer-bonded explosive (PBX) formulations were prepared via aqueous suspension coating: a reference system HMX/Viton A, known as a standardized explosive PBXN-5, and a modified system containing 1 wt.% IF-W_2 dispersed in the polymer solution. The binder was introduced as a solution in ethyl acetate, while nanoparticles were ultrasonically dispersed. Rheological behavior of the solutions were examined by viscosity measurements over time, performed using an MCR 302 rheometer at constant shear rate and temperature. The presence of IF-WS_2 significantly accelerated increase in viscosity, indicating enhanced solvent evaporation and possible nanoparticle–polymer interactions. After drying, PBX granules were characterized by bulk density, sieve analysis of the particle size distribution, microscopic analysis, DSC, and sensitivity to impact and friction. The results suggest that IF-WS_2 nanoparticles influence both processing behavior and final material properties, offering potential for improved control of PBX performance and safety.

Keywords: *HMX, Viton A, PBX, IF-WS_2 nanoparticles, rheology, sensitivity.*

1. INTRODUCTION

Polymer-bonded explosives (PBX) represent an important class of energetic materials designed to combine high explosive performance with improved mechanical properties and reduced sensitivity to accidental initiation. Their structure consists of energetic crystalline particles embedded in a polymeric binder matrix, which provides mechanical integrity, improves handling safety, and protects explosive crystals from external mechanical stimuli. They may be produced by different technologies and using various types of polymers: like granular formulations with thermoplastic binder, intended for pressing, then like cast-cured polymers with elastomeric binder, or like hand-mealable putty-like explosives with highly plasticized partly cured elastomeric binders [1-4]. In case of granular compositions they represent explosive materials in which explosive powder is bound together in a matrix using 5-10wt.% of a synthetic polymer. These PBXs were first developed in 1952 at Los Alamos National Laboratory, as RDX embedded in polystyrene with plasticizer [1, 5]. Later on, during the following decades, formulations were developed using other nitramines, like HMX, PETN, CL-20, or other high explosives, with a wide variety of polymer phlegmatizers, like polycarbonate, PMMA, Teflon, fluoropolymers like Viton, thermoplastic elastomers like Estane, and some energetic polymers [2, 4].

Corresponding author: D. Bajić;
E-mail: danica.bajic@mod.gov.rs
DOI: 10.5937/OTEH26047B

Among the high explosives used in PBXs, hexogen (RDX) and octogen (HMX) remain most widely employed secondary explosives due to their high detonation velocity, high detonation pressure, and excellent energy output. However, the sensitivity to impact and friction necessitates the application of efficient phlegmatization techniques to improve safety without significantly compromising energetic performance.

Fluoropolymer binders are frequently utilized in PBXs because of their excellent chemical compatibility with nitramines, thermal stability, hydrophobicity, and ability to reduce mechanical sensitivity [4-6]. Viton A, a copolymer based on vinylidene fluoride and hexafluoropropylene, is among the most commonly used fluoropolymer binders in military PBX formulations. One of the best-known examples is PBXN-5, which consists predominantly of HMX particles coated with a small amount of Viton A binder (95/5 wt.%). The performance and safety characteristics of such materials strongly depend on the quality of particle coating, binder distribution, particle size distribution, and the microstructure developed during the coating process [7-9].

In recent years, significant attention has been devoted to the incorporation of nanomaterials into energetic formulations. Various nanoparticles have been investigated as additives capable of modifying mechanical, thermal, and sensitivity characteristics of explosives and propellants. Due to their high specific surface area and unique physicochemical properties, nanoparticles can alter crystal-binder interactions, influence heat transfer mechanisms, modify rheological behavior during processing, and improve resistance to accidental mechanical initiation [10]. The use of nanostructured additives therefore offers new possibilities for tailoring the performance and safety of advanced energetic materials.

Among the available nanomaterials, inorganic fullerene-like tungsten disulfide (IF-WS_2) nanoparticles have attracted considerable scientific and technological interest. Unlike conventional layered WS_2 particles, IF-WS_2 nanoparticles possess a closed-cage, quasi-spherical morphology consisting of multiple concentric WS_2 layers. This unique structure provides exceptional tribological properties, including low friction coefficients, high wear resistance, and efficient energy dissipation under mechanical loading. These nanoparticles exhibit excellent thermal stability, high mechanical strength, and good compatibility with polymer matrices [11-18]. Previous studies have demonstrated their effectiveness as reinforcing and lubricating additives in various polymer composites, where improvements in mechanical and tribological performance have been observed [17-22]. The tribological characteristics of IF-WS_2 nanoparticles suggest that they may also contribute to reducing the sensitivity of energetic materials. Mechanical initiation of PBXs is commonly associated with localized energy concentration and hotspot formation occurring at crystal contacts, defects, or frictional interfaces. By reducing interparticle friction and promoting more uniform stress distribution, IF-WS_2 nanoparticles may mitigate the mechanisms responsible for hotspot generation. In addition, their presence may influence the rheological behavior of polymer binder solutions during PBX preparation, affecting coating efficiency, solvent evaporation, particle agglomeration, and the final microstructure of the explosive composition.

Despite the extensive investigation of IF-WS_2 nanoparticles in tribological and polymer-composite applications, their potential use in polymer-bonded explosives has not been explored or not reported. In particular, little information is available regarding their influence on the rheological properties of fluoropolymer binder systems and their role during the aqueous suspension coating process, which is commonly employed for RDX or HMX phlegmatization. Therefore, the objective of this study was to investigate the effect of IF-WS_2 nanoparticles on the preparation and properties of PBXN-5 explosive. Particular attention was devoted to the rheological behavior of Viton A solutions containing dispersed nanoparticles, as well as to the resulting particle morphology, granulometric characteristics, thermal behavior, and sensitivity to impact and friction.

2. EXPERIMENTAL

2.1. Materials and PBX preparation

The materials used for the preparation of PBX formulations in this study were the following:
- octogen (HMX), supplied by Prva Iskra – Namenska, Barič, Serbia;
- Viton A fluoropolymer binder (DuPont);
- inorganic fullerene-like tungsten disulfide nanoparticles (IF-WS_2) produced by ApNano, NanoLub – Israel;
- ethyl acetate as the solvent for the polymer binder, MERCK; and
- distilled water.

Two HMX-based PBX formulations were prepared. The reference composition corresponded to a conventional PBXN-5 formulation consisting of HMX coated with Viton A binder in ratio 95/5 wt.%. The modified formulation contained the same HMX/Viton A system with the addition of 1 wt.% IF-WS_2 nanoparticles (regarding the amount of Viton A) dispersed within the polymer solution prior to coating. The PBX samples were prepared using the aqueous suspension coating method based on polymer precipitation from solution. HMX crystals were first dispersed in water in a jacketed glass reactor equipped with mechanical stirring and temperature control. Viton A was dissolved in ethyl acetate to obtain a homogeneous polymer solution. For the nanomodified formulation, IF-WS_2 nanoparticles were dispersed in the Viton A solution by ultrasonic treatment before addition to the explosive suspension in order to ensure uniform nanoparticle distribution. The polymer solution was then gradually introduced into the stirred HMX/water suspension. During continuous stirring and controlled heating, ethyl acetate evaporated, causing precipitation of the polymer onto the surface of HMX crystals, producing PBX granules. In the nanomodified system, IF-WS_2 nanoparticles became incorporated within the precipitated

polymer coating.

After the complete solvent evaporation, the coated granules were separated, thoroughly washed with water to remove residual processing impurities, and dried in a laboratory oven at moderate temperature to avoid softening of the polymer binder and agglomeration of the granules. Part of the preparation procedure is schematically shown in Figure 1.

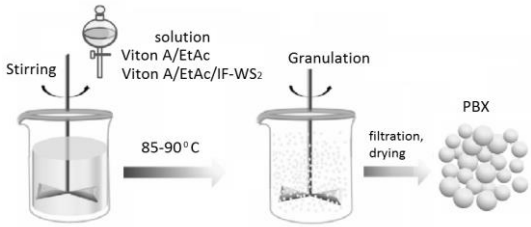


Figure 1. Ilustration of preparation of PBX samples

2.2. Methods

Prior to addition of the HMX suspension in water, small amount of the polymer phlegmatizer solutions in ethyl acetate, with and without the nanoparticles of tungsten disulfide, was taken aside and submitted to examination of viscosity on a rheometer, to monitor the change in apparent viscosity with time caused by the evaporation of ethyl acetate and to evaluate the influence of WS_2 particles on this change. The measurements were carried out using an Anton Paar MCR 302 rotational rheometer equipped with a 35 mm parallel-plate measuring system and P-PTD 220 and H-PTD 220 temperature-control elements. The gap between the plates was set to 1 mm, and the measurement temperature was maintained at 25 °C. Each sample was tested at a constant shear rate of 0.1 s^{-1} for 30 min. The apparent viscosity was continuously recorded as a function of time.

The prepared PBX granules were characterized using a combination of physical, chemical, thermal, and sensitivity analyses. Bulk density was determined by measuring the mass of a known volume of freely poured granules. The morphology and surface characteristics of the coated granules were examined by optical microscopy in order to assess particle shape, coating quality, and the presence of coarse agglomerates. This control of the phlegmatization success was done using LEICA M205A stereo microscope. Particle size distribution was evaluated by dynamic image analysis using DIA 500 Anton Paar device, in liquid mode, with water as a dispersion medium.

Thermal behavior was investigated by differential scanning calorimetry (DSC), using device Nexta DSC200, to determine characteristic phase transitions and decomposition temperatures of the prepared compositions and to assess the influence of IF-WS_2 nanoparticles on the thermal stability of the explosive material. Measurement was done in a temperature interval -80 to 325°C.

The safety characteristics of the PBX formulations were evaluated through standard sensitivity tests. Impact sensitivity was determined using a drop-hammer apparatus Julius Peters by measuring the response of the samples to controlled mechanical impact, using standard steel collars and cylinders made by OZM Research. Friction sensitivity was assessed using a friction tester Julius Peters under standardized loading conditions, using standard porcelain plates and pins produced by OZM Research.

3. RESULTS AND DISCUSION

3.1. Rheological behavior of the polymer solutions

The examined samples of neat and nanomodified Viton A/EA system showed a similar three-stage change in apparent viscosity with time, shown in Figure 2.

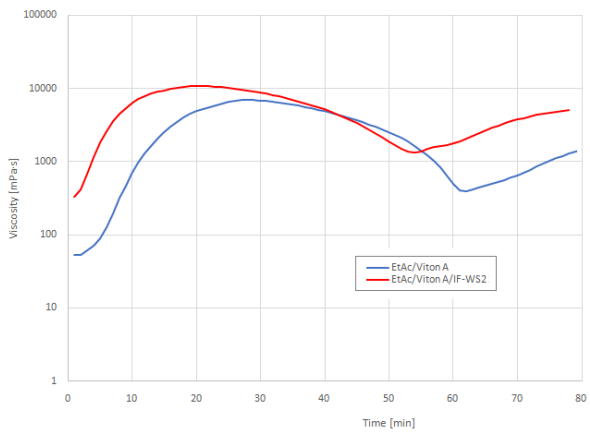


Figure 2. Apparent viscosity as a function of time for Viton A/ethyl acetate solutions with and without WS_2.

The apparent viscosity increased during the first 7-9 min of the measurement until a maximum value was reached. This initial increase was expected due to the progressive evaporation of ethyl acetate, which increased the concentration of Viton A and, in the nanomodified sample, the concentration of solid particles. After reaching the maximum, the apparent viscosity decreased until approximately 18-21 min. A second increase was then observed and continued until the end of the measurement. As it can be seen, a higher peak apparent viscosity was detected for the Viton A/EA/WS_2 sample, which reached 10.77 Pa·s after 7.0 min. Viton A/EA reference sample reached 7.09 Pa·s after 9.3 min. The apparent viscosity decreased to 1.32 Pa·s at 18.0 min for the WS_2-containing sample, and 0.393 Pa·s at 20.7 min for the sample without the nanoparticles. At the end of the measurement, the apparent viscosity increased again to 7.24 Pa·s for the Viton A/EA/IF-WS_2 dispersion, and 2.40 Pa·s for the Viton A/EA system. In the parallel-plate geometry, the sample was exposed to the surrounding atmosphere mainly along the outer edge of the plates. Therefore, solvent evaporation was probably more pronounced in this region, resulting in the formation of a solvent-depleted, polymer-rich layer. Under continuous shear, this layer may have been progressively disrupted or partially detached. Edge effects, including possible wall slip or partial loss of contact between the sample and the measuring plates, may have reduced the measured torque

and caused the observed decrease in apparent viscosity. The subsequent increase was most likely related to the continued evaporation of ethyl acetate and the resulting increase in the concentration of Viton A and solid particles in the remaining sample.

3.2. Morphology of the PBX granules

The appearance of the granules of the reference PBXN-5 and the nanomodified formulation are given in Figure 3. As may be observed, both formulations appear to have expected morphology of the granules, typical for this type of PBX [7-9].

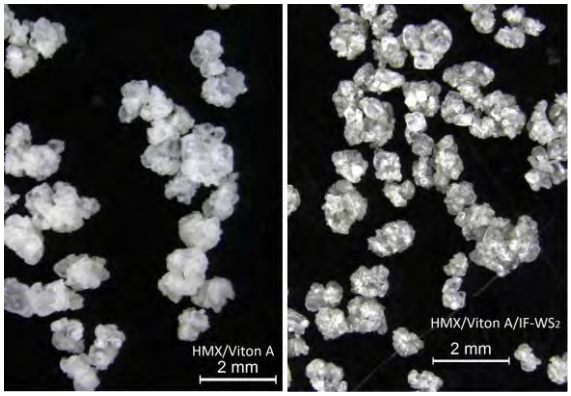


Figure 3. Micrographs of the PBX granules

3.3. FTIR analysis results

FTIR spectra registered for the analyzed PBXs are given in Figure 4.

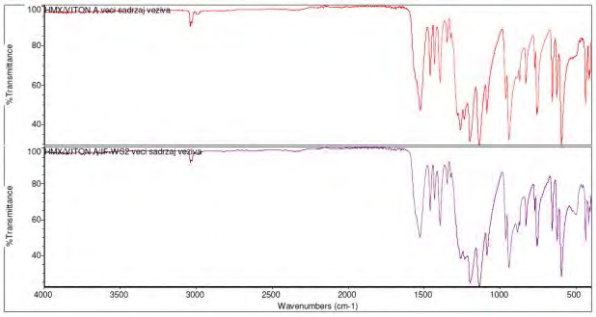


Figure 4. FTIR spectra of the prepared PBXs

It may be observed that there is no significant difference between the registered spectra, so the addition of WS_2 nanoparticles does not have a notable chemical interaction. FTIR spectra of the reference and nano-modified PBXN-5 samples exhibited the same characteristic absorption bands. No new absorption peaks or significant band shifts were observed after incorporation of IF-WS_2 nanoparticles, indicating that the modification did not induce chemical changes in the PBX matrix. Minor variations in band intensities suggest predominantly physical interactions between the nanoparticles and the composite constituents, and to differences in component distribution at the specific point of ATR contact.

3.4. Granulometric size distribution and loose bulk density of the PBX granules

The results of the measurement of the particle size are given in Figure 5. It may be noted that the obtained granules had practicaly same size distribution.

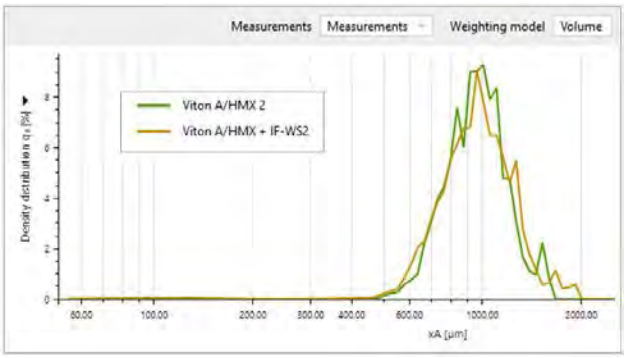


Figure 5. Size distribution of the PBX granules

For the reference PBXN-5 sample, loose bulk density was measured to be 654.1 g/dm^3, and for the WS_2-nanomodified PBXN-5 it was 667.1 g/dm^3. The obtained values are similar to those obtained earlier for PBXN-5 [7-9]. This small difference in the loose bulk density might have explanation in improved packing and reduced void volume caused by the presence of IF-WS_2 nanoparticles in the phlegmatizer [23-26].

3.4. DSC results

DSC thermograms of the HMX/Viton and HMX/Viton/IF-WS_2 formulations exhibited similar thermal behavior over the entire investigated temperature range. The DSC thermograph of the analyzed PBX sample HMX/Viton A is given in Figure 6.

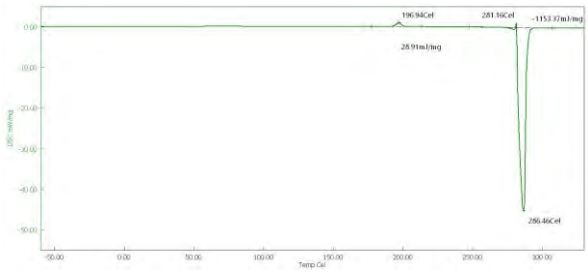


Figure 6. DSC thermograph for PBXN-5

An endothermic event observed at approximately 197-198°C corresponds to the well-known β to δ polymorphic transition of HMX. The transition temperature showed only a minor increase from 196.94 °C for HMX/Viton to 198.22 °C for HMX/Viton/IF-WS_2. At higher temperatures, both formulations displayed a pronounced exothermic decomposition peak centered at approximately 286.5 °C, indicating that the incorporation of IF-WS_2 nanoparticles does not significantly affect the thermal stability of HMX. A small increase in the decomposition peak temperature was observed for the nanomodified formulation (286.54 °C compared to 286.46 °C for the reference sample), although the magnitude of this shift is within the typical uncertainty of DSC measurements. Similarly, the melting endotherm of

HMX, partially overlapped by the onset of decomposition, appeared at slightly higher temperatures in the presence of IF-WS_2 (281.50 °C compared to 281.16 °C). Overall, all characteristic thermal events exhibited a consistent but very small shift toward higher temperatures after the addition of IF-WS_2, suggesting a weak stabilizing effect associated with nanoparticle-matrix interactions. However, additional measurements would be required to confirm the statistical significance of this trend. The slightly higher decomposition enthalpy observed for the formulation with IF-WS_2 may indicate a more complete decomposition process resulting from improved thermal coupling and enhanced interfacial contact within the composite microstructure. A higher decomposition enthalpy was recorded for the IF-WS_2-modified formulation (1263 J g^{-1}) compared to the reference composition (1153 J g^{-1}). However, considering that only a single DSC measurement was performed for each sample and that the analyzed sample masses were approximately 1 mg, this difference should be interpreted cautiously until confirmed by replicate measurements.

A more pronounced effect of IF-WS_2 was observed in the glass transition region of the Viton binder. The part of DSC thermographs of the compared samples where we may observe enlarged part with detectable Tg is given in Figure 7.

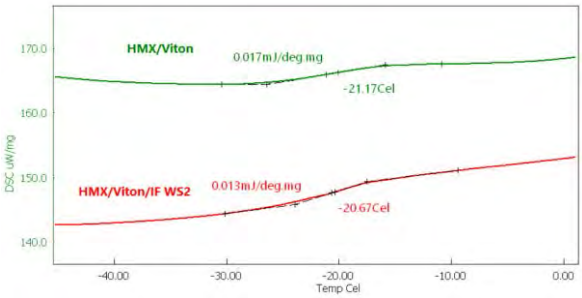


Figure 7. Enlarged DSC thermographs with Tg of Viton A detected in PBXN-5 and PBXN-5/IF-WS_2

in the glass transition region revealed a Tg value of -21.17 °C for the HMX/Viton formulation and -20.67 °C for the HMX/Viton/IF-WS_2 composite. The incorporation of IF-WS_2 nanoparticles therefore resulted in a slight increase of the glass transition temperature by approximately 0.5 °C. Although the shift in Tg is relatively small, a noticeable reduction in the heat capacity increment at the transition was observed. The ΔCp value decreased from 0.017 $mJ \cdot deg^{-1} \cdot mg^{-1}$ for the reference formulation to 0.013 $mJ \cdot deg^{-1} \cdot mg^{-1}$ after the addition of IF-WS_2, corresponding to a reduction of approximately 24%. The negligible change in Tg suggests that IF-WS_2 does not significantly alter the molecular structure of the Viton binder nor induce extensive changes in chain flexibility. However, the reduced ΔCp indicates the presence of restricted polymer segments at the polymer–nanoparticle interface. Such behavior is commonly attributed to physical adsorption of polymer chains on the nanoparticle surface, resulting in the formation of an interfacial region with reduced segmental mobility. The nanoparticles slightly constrain the motion of Viton chains without substantially affecting the overall glass transition temperature. The reduced ΔCp indicates that a fraction of the Viton chains becomes immobilized at the IF-WS_2 surface, suggesting stronger filler-binder interactions and a more constrained interfacial polymer layer.

3.5. Sensitivity to impact and friction

The results of sensitivity tests performed on the prepared granular PBXs are given in Table 1.

Table 1. Sensitivity of the prepared granular PBXs

Sample	Impact sensitivity	Friction sensitivity
PBXN-5	3.14 J	324 N
PBXN-5/IF-WS_2	2.94 J	240 N
PBXN-5/graphite	2.76 J	268 N

Contrary to the expected desensitizing effect of solid lubricant nanoparticles, IF-WS_2 slightly increased the mechanical sensitivity of PBXN-5. This behavior may be attributed to the introduction of additional interfacial regions, local stress concentration sites within the PBX microstructure and to the presence of hard nanoscale particles. This might have helped formation of hot-spot under mechanical stimuli [6, 23]. This means that in order to achieve the desired effects, different approach of incorporation of the nanostructures should be applied, so the further research will consider possibility of applying a thin layer of the solid lubricating particles ontop of the PBX granules.

CONCLUSIONS

Nanomodified PBXN-5 was prepared and characterized. The incorporation of 1 wt.% IF-WS_2 nanoparticles into the Viton A binder influenced the processing behavior and properties of PBXN-5 explosive. Rheological measurements showed a significant increase in apparent viscosity of the Viton A/ethyl acetate solution with added nanoparticles during solvent evaporation, indicating interactions between the nanoparticles and the polymer solution that may affect coating efficiency and granule formation. FTIR analyses confirmed that the addition of IF-WS_2 did not induce chemical changes in the PBX matrix. DSC demonstrated that the presence of the nanoparticles slightly altered PBX`s thermal stability, with a slight increase in Tg and reduction in ΔCp. The nanomodified formulation exhibited morphology comparable to the reference PBXN-5. However, a slight increase in mechanical sensitivity was observed, which will direct further research towards different technology of incorporation of IF-WS_2 nanoparticles in formulation of PBXs.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325, and University of Defense, Military Academy, Proj. No. VA-TT/1/22-24.

References

[1] GIBBS, T.R., POPOLATO, A.: LASL Explosive Property Data, University of California Press, 1980.

[2] AGRAWAL, J.P.: High Energy Materials:

Propellants, Explosives and Pyrotechnics, 2010.

[3] MEYER, R., KÖHLER, J., HOMBURG, A.: Explosives, 6th Ed., Wiley-VCH, Weinheim, 2007.

[4] SIMIC D: Liveni termobarični PBX eksplozivi, Vojnotehnički institut Beograd, Naučnotehničke informacije, 53, 2, ISSN 1820-3418, 2016.

[5] AKHAVAN, J.: The Chemistry of Explosives, 2nd Ed., Royal Society of Chemistry, Cambridge, 2004.

[6] DAS, P., UDAYKUMAR, H.S.: Mechanisms of shock-induced initiation at micro-scale defects in energetic crystal-binder systems, Shock Waves, 32 (2022), 593–616.

[7] TERZIĆ, S.: Influence of Coating Techniques and Organic Solvents on the Quality of Granulated PBX Based on HMX, Scientific Technical Review, 52(1) (2007), 34–38.

[8] TERZIĆ, S.: Tehnološki postupci dobijanja granulisanih PBX na bazi oktogena i Vitona A, int. dok., VTI-04-01-0394, Vojnotehnički institut, Beograd, 2004.

[9] TERZIĆ, S.: Optimizacija tehnološkog postupka izrade granulisanih PBX na bazi oktogena i Vitona A, int. dok., VTI-04-01-0436, Vojnotehnički institut, Beograd, 2005.

[10] X. Wang, P.Wan et al. Efficiently enhancing thermal conductivity of polymer bonded explosives via the construction of primary-secondary thermal conductivity networks, Defence Technology 48 (2025) 95-103.

[11] TENNE, R., MARGULIS, L., GENUT, M., HODES, G.: Polyhedral and Cylindrical Structures of Tungsten Disulfide, Nature, 360 (1992), 444–446.

[12] TEVET, O., VON-HUTH, P., et al.: Friction Mechanism of Individual Multi-Layered Nanoparticles, Proceedings of the National Academy of Sciences of the United States of America, 108(50) (2011), 19901–19906.

[13] RAPOPORT, L., TENNE, R. et al.: Tribological Applications of WS2 and MoS2 Inorganic Fullerene-Like Nanoparticles as Solid Lubricants, Wear, 225–229 (1999), 975–982.

[14] RAPOPORT, L., NEPOMNYASHCHY, O. et al.: Behaviour of fullerene-like WS_2 nanoparticles under severe contact conditions, Wear, 259(1–6) (2005), 703–707.

[15] ABATE, F., D'AGOSTINO, et al.: Tribological Behaviour of MoS2 and Inorganic Fullerene-Like WS_2 Nanoparticles Under Boundary and Mixed Lubrication Regimes, Tribology in Industry, 32(4) (2010), 91–98.

[16] RAPOPORT, L., BILIK, et al.: Hollow Nanoparticles of WS2 as Potential Solid-State Lubricants, Nature, 387 (1997), 791–793.

[17] NAFFAKH, M., DÍEZ-PASCUAL, A.M., MARCO, C., et al.: Novel Melt-Processable PEEK/IF-WS2 Nanocomposites for Critical Applications, Journal of Physical Chemistry B, 114 (2010), 11444–11453.

[18] HAZARIKA, S.J., MOHANTA, D.: Revealing Mechanical, Tribological, and Surface-Wettability Features of Nanoscale Inorganic Fullerene-Type Tungsten Disulfide Dispersed in a Polymer, Journal of Materials Research, 34(21) (2019), 3666–3677.

[19] SETHULEKSHMI, A., JAYAN, J.S., SARITHA, A., JOSEPH, K.: Insights into the Reinforcibility and Multifarious Role of WS2 in Polymer Matrix, Journal of Alloys and Compounds, 876 (2021), 160107.

[20] NAFFAKH, M., MARTÍN et al.: Influence of Inorganic Fullerene-Like WS2 Nanoparticles on the Thermal Behavior of Isotactic Polypropylene, Journal of Polymer Science Part B: Polymer Physics, 45(17) (2007), 2309–2321.

[21] M. MARJANOVIĆ, D. SIMIĆ et al. Carbon-Epoxy Composites Reinforced with Nano-Structures of Tungsten Disulfide for Potential Use in Aircraft Structures, Scientific Technical Review, 68,3, 2018, 13-17.

[22] SIMIĆ, D. M., STOJANOVIĆ, D. B., DIMIĆ, M., MIŠKOVIĆ, K., MARJANOVIĆ, M., BURZIĆ, Z., USKOKOVIĆ, P. S., ZAK, A., TENNE, R. (2019). Impact resistant hybrid composites reinforced with inorganic nanoparticles and nanotubes of WS2. Composites Part B: Engineering, 176, 107222.

[23] LONG, Y., LIU, Y., NIE, F., & CHEN, J. (2012). Theoretical study of breaking and slipping processes for HMX/graphite interface. Applied Surface Science, 258(7),2384-2392. https://doi.org/10.1016/j.apsusc.2011.10.052

[24] CHUN-LEI JIANG et al., A review on packing characteristics of explosives: Recent advances and future perspectives, Energetic Materials Frontiers, https://doi.org/10.1016/j.enmf.2026.04.008

[25] YU, J., XU, S., PANG, W. et al. (2024). The Structural Regulation and Properties of Energetic Materials: A Review. Nanomaterials, 15(15), 1140. https://doi.org/10.3390/nano15151140

[26] QIN, Y., SONG, Z et al. (2025). Research Progress on Mechanical Enhancement of Polymer-Based Energetic Composites by Coating of the Crystalline Fillers. Polymer Composites, 46(17), 15535-15566. https://doi.org/10.1002/pc.70002

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

APPLICATION OF A CASCADE OF NEURAL NETWORKS FOR MODELING AND PREDICTING DETONATION CHARACTERISTICS OF HIGH-ENERGY MATERIALS

V.S. ABRUKOV, D.A. ANUFRIEVA, S.S. KUZMIN, D.A. SAPEROV

I.N. Ulyanov Chuvash State University, Cheboksary, Russia, abrukov@yandex.ru

Abstract: *Machine learning methods, particularly neural networks, offer the ability to identify complex nonlinear relationships between detonation characteristics, structure, and composition of HEMs. This work employs a novel approach—a cascade of neural networks—to improve model quality and enable comprehensive analysis of HEM characteristics.*

Keywords: *Machine learning, neural networks,high-energy materials, detonation characteristics.*

1. INTRODUCTION

Modern challenges in the study of high-energy materials (HEMs) require effective methods for predicting performance characteristics.
Traditional approaches based on experimental studies and theoretical calculations are often resource-intensive.

Machine learning methods, particularly neural networks (NNs), offer the ability to identify complex nonlinear relationships between detonation characteristics, structure, and composition of HEMs [1-3].

This work employs a novel approach—a cascade of neural networks—to improve model quality and enable comprehensive analysis of HEM characteristics.

2. MATERIALS AND METHODS

The study was conducted on a dataset comprising 32 individual and composite explosives. For each HEM, 7 physicochemical parameters were considered: density, oxygen balance, nitrogen content, total nitrogen and oxygen content, standard enthalpy of formation, melting point, and decomposition point. The Table of experimental results used is depicted in the end of paper.

A cascade neural network method (sequential approximation) was used to predict four target variables: detonation velocity, detonation pressure, impact sensitivity, and friction sensitivity. The method proceeds as follows:

The first neural network (NN1) is trained directly on experimental data: 7 HEM descriptors are input, and one detonation characteristic is output. The values obtained from NN1 are used as output targets for training the second neural network (NN2). The third cascade level (NN3) is trained on the output data of NN2, resulting in a maximally smooth dependence. Each subsequent network uses the previous network's output as target values, equivalent to sequential filtering of high-frequency noise. The final model produces a smooth mapping with RMSE below 10^{-5}.

For objective validation, the dataset was divided into training (80%) and test (20%) sets. Training was stopped when the minimum validation error was reached. Five-fold cross-validation confirmed stability (mean R^2 for detonation velocity: 0.994±0.003).

3. RESULTS

The constructed models demonstrate high quality (Table 1).

Table 1. Model quality metrics

Parameter	R^2	MSE	MAE	MAPE %
Detonation velocity	0.996	0.070	0.025	0.76
Detonation pressure	0.993	0.662	0.002	0.27
Impact sensitivity	0.862	5.611	1.743	22.37
Friction sensitivity	0.877	42.35	8.624	6.92

Comparison with alternative methods [4-6] shows typical R^2 values of ~0.95-0.99, while our cascade approach achieves $R^2 \approx 0.996$ and 0.993 for detonation velocity and pressure, confirming its high quality.

Analysis of physical regularities: The models reveal fundamental explosion physics: monotonic increase of detonation velocity and pressure with density (following classical detonation equations [7, 8]), dependence of detonation velocity on enthalpy of formation, and dependence of impact sensitivity on nitrogen content and decomposition temperature.

Corresponding author: V.S. Abrukov;
E-mail: abrukov@yandex.ru

DOI: 10.5937/OTEH26048A

Some of the results are depicted in Figures 1-5.

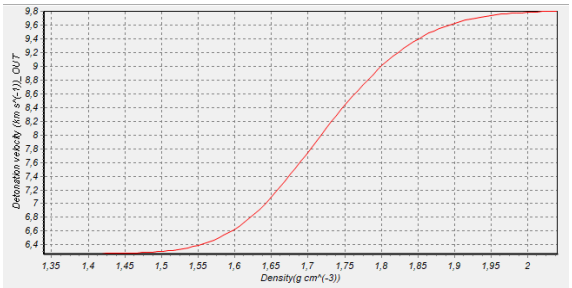


Figure 1. Dependence of detonation velocity on the density of the explosive

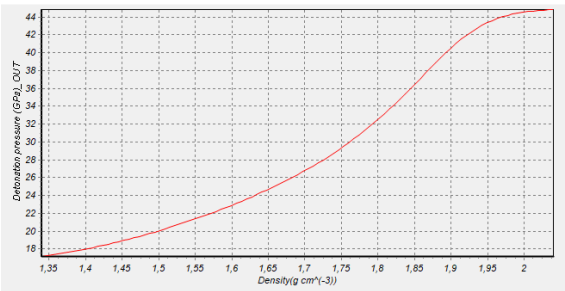


Figure 2. Dependence of detonation pressure on the density of the explosive

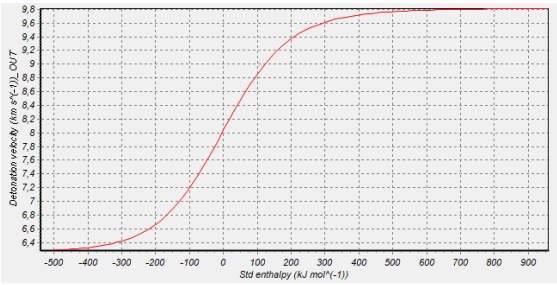


Figure 3. Dependence of detonation velocity on standard enthalpy of formation

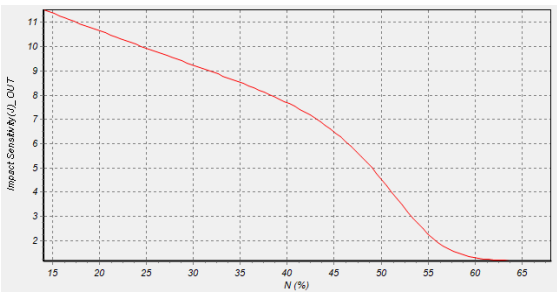


Figure 4. Impact sensitivity as a function of nitrogen content

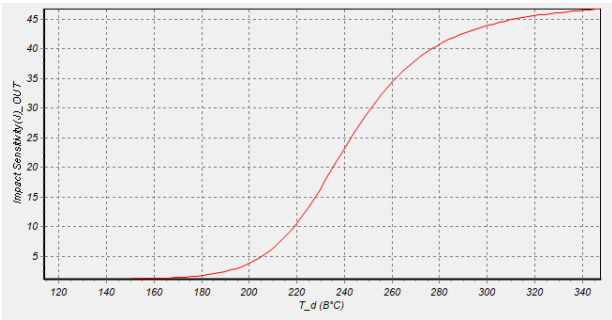


Figure 5. Impact sensitivity as a function of decomposition temperature

Only the most common dependencies are shown. Analysis shows maximum influence on detonation velocity/pressure from density and oxygen balance, while sensitivity depends most on total nitrogen and oxygen content and decomposition temperature, agreeing with physicochemical concepts.

4. DISCUSSION

While only the most common dependencies are visualized in Figures 1–5, a deeper analysis of the constructed cascade neural network models reveals a strong alignment with classical detonation physics and thermochemistry.

Regarding detonation performance, the models confirm the paramount influence of density (ρ) and oxygen balance (Ω_CO) on detonation velocity (D) and pressure (P).

As illustrated in Figures 1 and 2, there is a pronounced monotonic positive correlation between density and both D and P. This is physically governed by the Chapman-Jouguet theory, where higher density implies tighter molecular packing, leading to a greater concentration of reactive sites and higher reaction zone pressures. For instance, the model accurately maps the transition from low-density DNAN (ρ = 1.34 g/cm³, D = 6.26 km/s) to high-density CL-20 (ρ = 2.04 g/cm³, D = 9.85 km/s).

Furthermore, Figure 3 demonstrates the critical dependence of detonation velocity on the standard enthalpy of formation (ΔH_f). Compounds with highly positive (endothermic) enthalpies of formation, such as CL-20 (397 kJ/mol) or BNCP (350 kJ/mol), store substantial chemical energy that is released upon decomposition, significantly boosting detonation velocity compared to exothermic counterparts like TNT or PETN, assuming comparable densities. Oxygen balance also plays a vital role; values closer to zero optimize the formation of stable gaseous products (CO_2 and H_2O), maximizing the heat of explosion and, consequently, D and P.

Conversely, the safety and sensitivity characteristics—impact sensitivity (IS) and friction sensitivity (FS)—are governed by different structural and thermodynamic factors.

The models indicate that sensitivity depends most heavily on the total nitrogen and oxygen content (N+O) and the decomposition temperature (T_d).

Figure 4 highlights a distinct inverse relationship between nitrogen content and impact safety: high-nitrogen

compounds (e.g., $C_4N_{10}O$ with 68.6% N) possess numerous relatively weak N–N and N–O bonds and high positive heats of formation, resulting in extreme sensitivity (IS = 0.9 J). In contrast, lower nitrogen content generally correlates with higher structural stability.

Additionally, Figure 5 confirms a strong direct correlation between thermal stability and mechanical insensitivity. Materials with high decomposition temperatures, such as TATB (T_d = 350 °C) and TNT (T_d = 300 °C), possess high activation energies for thermal decomposition, which translates to a higher threshold for mechanical initiation (high IS values of 50 J and 39.2 J, respectively).

Ultimately, the cascade neural network successfully extracted these complex, non-linear physicochemical regularities from the dataset. The fact that the model independently converged on dependencies that perfectly agree with established theoretical concepts validates its reliability for the comprehensive analysis and predictive design of high-energy materials.

Future work involves database expansion and creating a "High-Energy Materials Genome" [9-13].

References

[1] Mukhutdinov A.R. et al. Information Technologies and Computing Systems, 2021, no. 4, pp. 83-90.
[2] Perederin Yu.V., Popok N.I. Chemical Physics and Mesoscopy, 2019, vol. 21, no. 2, pp. 245-253.
[3] Lyasheva S.A. et al. Bulletin of Technological University, 2015, vol. 18, no. 15, pp. 226-229.
[4] Elton D. et al. Scientific Reports, 2018, vol. 8, 9059.
[5] Elton D. et al. Molecular Systems Design & Engineering, 2019, vol. 4, pp. 828-849.
[6] Liu W.H. et al. Journal of Energy Chemistry, 2025, vol. 102, pp. 161-171.
[7] Zeldovich Ya.B., Raizer Yu.P. Physics of Shock Waves, Nauka, 1966.
[8] Fickett W., Davis W.C. Detonation: Theory and Experiment, Dover, 2000.
[9] Abrukov S.V. et al. International Journal of Energetic Materials and Chemical Propulsion, 2010, vol. 9, no. 5.
[10] Chandrasekaran N., Abrukov S.V. et al. Propellants, Explosives, Pyrotechnics, 2019, vol. 44, no. 5, pp. 579-587.
[11] Pang W., Abrukov S.V. et al. Crystals, 2023, vol. 13, no. 2, pp. 237-252.
[12]. Doherty R.M., Abrukov V.S., Boukouvalas Z., Baek S., Terreux R. Application of AI and ML to Novel Energetic Materials // Energetic Materials and Techniques: Advances in Chemical Propulsion and Power Generation : in 2 vol. / ed. by L.T. DeLuca, R.M. Doherty. — Wiley, 2026. — Chapter 1. — ISBN 978-3-527-35545-7. — https://doi.org/10.1002/9783527853267.ch01
[13]. Abrukov V., Pang W., Anufrieva D., Sapyorov D. Neural Networks in the Development of High-Energy Materials Genome // Energetic Materials and Techniques: Advances in Chemical Propulsion and Power Generation : in 2 vol. / ed. by L.T. DeLuca, R.M. Doherty. — Wiley, 2026. — Chapter 3. — ISBN 978-3-527-35545-7. — https://doi.org/10.1002/9783527853267.ch03

Table. Experimental results.

Formula or Name	IS (J)	FS (N)	D (km s^{-1})	P (GPa)	Additive	ρ (g cm^{-3})	Ω_CO (%)	N (%)	N+O (%)	ΔH_f (kJ mol^{-1})	T_m (°C)	T_d (°C)
1.C4N10O	0,9	5	8,5	32	None	1,75	-23,5	68,6	76,5	959	81	114
2.C4H2N8O	3	140	7,9	28	None	1,71	-36	62,9	71,9	574	314	179
3.C5H3N7O	50	360	7,1	24	None	1,68	-49,7	55,4	64,4	418	291	314
4.C6H2F3N7O	50	360	6,6	23	None	1,84	-39,2	40	46,5	-220	152	314
5.C4HN7O2	5	100	7,9	29	None	1,79	-22,3	54,8	72,6	405	32	152
6.C5H2N6O2	17	360	7,1	24	None	1,75	-36	47,2	65,2	243	328	320
7.HNS	20	360	7,2	26	None	1,75	-17,8	18,7	61,3	68	320	320
8.TATB	50	353	7,91	29,7	None	1,94	-56	32,6	67,4	-139,5	350	350
9.HMX	7,4	120	9,11	39	None	1,91	-21,6	37,8	81,9	75	279	279
10.HMX	7,4	120	9,11	39	1% C60-1	1,91	-21,6	37,8	81,9	75	279	279
11.HMX	7,4	120	9,11	39	1% C60-2	1,91	-21,6	37,8	81,9	75	279	279
12.HMX	7,4	120	9,11	39	1% C60-3	1,91	-21,6	37,8	81,9	75	279	279
13.RDX	7,5	120	8,75	34,7	None	1,82	-21,6	37,8	81,9	70,3	210	210
14.CL-20	4	108	9,85	45	None	2,04	-10,9	38,4	77,2	397	240	240
15.e-CL-20	4	108	9,85	45	None	2,04	-10,9	38,4	77,2	397	240	240
16.e-CL-20	4	108	9,85	45	glue	2,04	-10,9	38,4	77,2	397	240	240

17.e-CL20	4	108	9,85	45	glue/graphene	2,04	-10,9	38,4	77,2	397	240	240
18.e-CL-20	4	108	9,85	45	glue/GO	2,04	-10,9	38,4	77,2	397	240	240
19.e-CL-20	4	108	9,85	45	glue/rGO	2,04	-10,9	38,4	77,2	397	240	240
20.TNT	39,2	353	6,82	19,5	None	1,65	-74	18,5	62,2	-59,3	80,9	300
21.PETN	7,4	60	8,4	33,5	None	1,77	-10,1	17,7	80,7	-538,7	141,3	190
22.PETN	7,4	60	8,4	33,5	1% CB	1,77	-10,1	17,7	80,7	-538,7	141,3	190
23.PETN	7,4	60	8,4	33,5	1% CNTs	1,77	-10,1	17,7	80,7	-538,7	141,3	190
24.FOX-7	20	252	8,87	34,4	None	1,88	-21,6	37,8	67,6	-133,9	240	240
25.ADN	15	120	8,3	25	None	1,82	25,9	45	77,8	-149	92	170
26.TKX-50	20	360	9,69	42,4	None	1,88	-15,9	58,6	74,4	251,5	221	221
27.DNAN	30	360	6,26	15,9	None	1,34	-95,2	14,1	52,9	-184	94,5	289
28.BNCP	10	240	8,9	35	None	1,9	-15	45,2	72,5	350	320	320
29.BNCP	12	260	8,95	35,5	MWCNTs-1	1,91	-15	45	72,3	348	322	322
30.BNCP	15	280	9	36	MWCNTs-2	1,92	-14,8	44,8	72,1	346	324	324
31.BNCP	18	300	9,05	36,5	SWCNTs	1,93	-14,5	44,5	71,8	344	326	326
32.BNCP	8	220	8,85	34,5	CB (Carbon Black)	1,89	-15,2	45,3	72,6	352	318	318

Note to Table.

Designations: IS (J) – impact sensitivity (J). FS (N) – friction sensitivity (N). D ($km \cdot s^{-1}$) – detonation velocity (km/s). P (GPa) – detonation pressure (GPa). Additive – type of additive. ρ ($g \cdot cm^{-3}$) – density. Ω_CO (%) – oxygen balance. N (%) – nitrogen content. N+O (%) – total nitrogen and oxygen content. ΔH_f ($kJ \cdot mol^{-1}$) – standard enthalpy of formation. T_m (°C) – melting point. T_d (°C) – decomposition temperature.

Explanation of abbreviations.

HNS – 2,2',4,4',6,6'-hexanitrostilbene. TATB – 1,3,5-triamino-2,4,6-trinitrobenzene. HMX – octogen (cyclotetramethylenetetranitramine). RDX – hexogen (cyclotrimethylenetrinitramine). CL-20 – hexanitrohexaazaisowurtzitane / ε-CL-20 – hexanitrohexaazaisowurtzitane with a modified crystal structure. TNT – 2,4,6-trinitrotoluene. PETN – pentaerythritol tetranitrate. FOX-7 – 1,1-diamino-2,2-dinitroethylene. ADN – ammonium dinitramide. TKX-50 – dihydroxylammonium 5,5'-bitetrazole-1,1'-dioxide. DNAN – 2,4-dinitroanisole. BNCP – bis(5-nitrotetrazolato-N2)tetraaminecobalt(III) perchlorate. C60-1, C60-2, C60-3 – various modifications of fullerene C60. CB (Carbon Black) – carbon black. CNTs – carbon nanotubes. MWCNTs-1, MWCNTs-2 – multi-walled carbon nanotubes. SWCNTs – single-walled carbon nanotubes. GO – graphene oxide. rGO – reduced graphene oxide. The compounds C4N10O, C4H2N8O, C5H3N7O, C6H2F3N7O, C4HN7O2, C5H2N6O2 do not have generally accepted names, therefore they are given as chemical formulas. For TKX-50 (No. 26), the table shows values at the theoretical maximum density.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

NUMERICAL–EXPERIMENTAL ANALYSIS OF EXPLOSIVELY FORMED PROJECTILE PENETRATION CHARACTERISTICS

ANDJELA IVKOVIĆ[1,2]
[1]Military Technical Institute, Belgrade, [2]Faculty of Engineering University of Kragujevac, andjamit99@gmail.com

SASA SAVIĆ
Military Technical Institute, Belgrade, sasa.savic@mod.gov.rs

JELENA MARINKOVIĆ
Military Technical Institute, Belgrade, jelena.marinkovic@vti.vs.rs

MLADEN TIMOTIJEVIĆ
Military Technical Institute, Belgrade, mladen.timotijević@mod.gov.rs

***Abstract**: Within this paper, a comparative analysis of the results obtained by numerical simulation and experimental testing of the penetration of an explosively formed projectile (EFP) was performed. The mentioned type of projectile is characterized by a complex process of penetrator formation. Understanding the dynamics of penetrator formation and its interaction with the target material is of key importance for analyzing the performance and comparing the construction parameters of these systems. The numerical model was developed with the aim of simulating the detonation process, projectile formation and studying its penetration. Experimental tests were conducted under strictly controlled conditions to determine the basic penetration parameters and validate the numerical model. The analysis includes a comparison of key parameters, including the geometry of the explosively formed projectile, the speed of movement and the depth of penetration in the target material. The obtained results indicate a significant agreement between the numerical and experimental data, thus confirming the adequacy of the applied numerical approach. Observed deviations were analyzed in the context of the simplification of the numerical model and limitations of the experimental methodology. The research results contribute to a better understanding of the mechanism of explosively formed projectile penetration and confirm the importance of numerical simulations as a reliable tool for evaluating their performance.*

***Keywords:** explosively formed projectile, numerical simulation, experimental testing, penetration, model validation.*

1. INTRODUCTION

An explosively formed projectile (EFP) is a specific type of warhead in which a metal casing is transformed into a compact, high-velocity projectile under the action of a detonation wave [1,2]. Unlike a conventional cumulative shaped charge, in which the metallic liner is typically conical and its collapse produces a high-velocity cumulative jet, an EFP (Explosively Formed Projectile) uses a shallow concave metallic liner, typically with a spherical or similar profile [1,2]. Under the influence of the detonation products, the EFP liner undergoes plastic deformation and forms a compact penetrator. Unlike a cumulative jet, which achieves its highest effectiveness within a specific and relatively limited range of stand-off distances and gradually loses coherence during flight, an EFP penetrator can maintain its integrity over significantly greater distances and achieve substantial penetration at longer stand-off distances from the target.

Its penetration characteristics depend not only on the velocity and mass of the projectile, but also on the penetrator geometry, mass distribution, stability in flight, and mechanical properties of the target material [3, 4]. The development of EFP research over the past few decades has been accompanied by advances in experimental techniques and numerical methods. Early research showed that projectiles with similar velocities can achieve different penetration depths depending on the projectile shape and characteristics of the target material, confirming that their performance cannot be reliably assessed based on a single ballistic parameter [3]. The development of three-dimensional numerical models has enabled the analysis of the complete process, from the formation and flight of the projectile to its interaction with the target. Comparison of numerical and experimental results has shown that reliable validation requires simultaneous analysis of the projectile geometry, its velocity, and the achieved penetration depth [4, 5].

Particular attention in modern research is paid to the influence of structural and material parameters on the process of formation and penetration of EFPs. The results show that the choice of liner material, its geometry and thickness distribution significantly affect the final shape, stability and penetration characteristics of the projectile [6, 7]. In addition, by changing the configuration of the liner, using a two-layer liner and detonation wave deflectors, it is possible to control the formation process and increase the length-to-diameter ratio of the penetrator [8]. Research on

Corresponding author: A.Ivković
E-mail: andjamit99@gmail.com

DOI: 10.5937/OTEH26049I

sintered liners has further indicated the importance of microstructure and manufacturing methods for preserving the integrity of the projectile during formation [9].

Modern approaches increasingly combine numerical simulations and experimental tests to more reliably assess the behavior of the projectile and the target. Experimental and numerical studies of penetration into different target materials have shown that measuring the penetration depth alone is not sufficient for a complete model assessment, but it is also necessary to analyze the geometry of the projectile and the characteristics of the resulting damage [10]. Validated numerical models are also used for design optimization, including the geometry of the cladding, projectile stabilization, and material selection, which significantly reduces the number of required experimental tests [11].

Based on the analyzed literature, it can be concluded that there is a need for further validation of numerical models through the simultaneous comparison of several characteristic parameters and different target configurations.

The aim of this paper is a comparative analysis of the results obtained by numerical simulation and experimental testing of the penetration of an explosively formed projectile. A numerical model was developed to simulate the detonation process, projectile formation and its interaction with the target material, while experimental tests were conducted under controlled conditions to determine the basic penetration parameters. By comparing the geometry of the formed projectile, its velocity and penetration depth, an analysis of the agreement between the numerical and experimental results was performed.

2. EXPERIMENTAL

In the following, preliminary warhead model has been designed and for it, the ANSYS Autodyn is used to simulate its EFP formation process. Used materials for each part of the EFP warhead and steel target are listed in Table 1.

Table 1. Materials used in numerical simulation

Component		Material	model	equation of state
Warhead-EFP	explosive	Octogen (HMX)	-	JWL
	liner (disc)	Cu-OFHC (Oxygen-Free High-Conductivity Copper)	Johnso-Cook	-
	case	EN AW-6061T6	Steinberg Guinan	-
	deviator	Polyethylene (PE)	-	-
Steel target		Armor steel (hardness up to 340HB)	Johnson-Cook	-

In accordance with the results obtained by numerical simulation, warheads were manufactured. All the components of the polygon tests (warhead, steel target) were in agreement with the numerical model. Used materials in accordance with the respective standards are listed in Table 2.

Table 2. Materials of components for the experimental set up

Component		Material
Warhead – EFP	explosive	phlegmatized octogen (HMX)
	liner (disc)	phosphorus-deoxidized copper (CW024A)
	case	EN AW-6061 T6
	deviator	Polyethylene (PE)
Steel target		Armor steel (hardness up to 340HB)

2.1 Numerical simulation

The numerical simulation applied in this research was developed with the aim of analyzing the formation process of an explosively formed projectile and its interaction with the target material (armour steel target). The simulation includes the detonation process of the explosive charge, the deformation of the liner (disk), the formation of the projectile, its movement through the air space and penetration into the steel target. The obtained results represent the basis for comparison with experimental tests and assessment of the ability of the numerical model to faithfully reproduce the physical processes that determine the performance of the EFP system [3, 4].

Due to pronounced deformations, high deformation rates and short duration of the process, the numerical model is based on the combined Euler–Lagrange approach. In this approach, the explosive charge and airspace are modeled in the Eulerian domain, while the liner (disc), warhead shell, and steel target are modeled in the Lagrangian domain. Such a formulation enables efficient modeling of the movement of detonation products while at the same time a precise description of the deformation of solid bodies and their mutual interaction, which is why it represents a standard approach in simulations of explosive processes [5, 10].

2.1.1 The geometry of EFP warhead model

The geometry of the numerical model of the warhead is defined in accordance with the geometric characteristics of the real experimental sample in order to ensure identical conditions for numerical simulation and experimental tests. Figure 1 shows a schematic representation of the geometry of the EFP warhead.

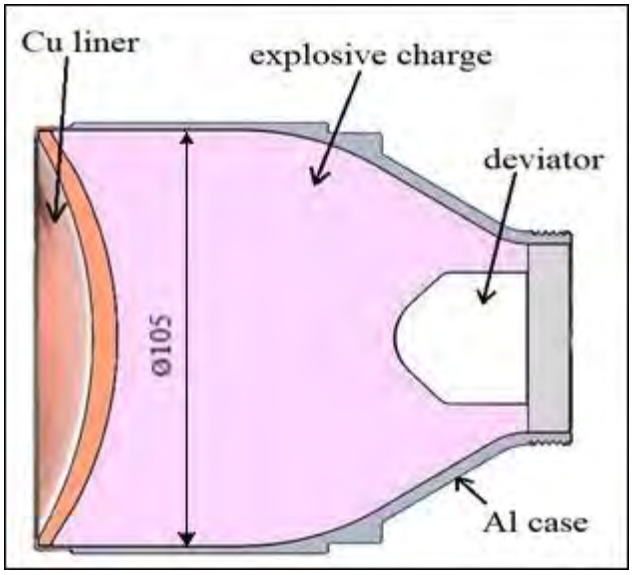


Figure 1. Geometric shape of the EFP warhead

2.1.2 Numerical models for description of EFP warhead elements and penetration process

To describe the material behavior of EFP warhead components, constitutive models and equations of state were used, which enable the simulation of processes characteristic of explosive loads, including large plastic deformations, high deformation rates and the spread of detonation products [3, 6].

The behavior of metal components is modeled using the Johnson–Cook constitutive model, which describes the dependence of yield stress on plastic strain, strain rate, and temperature. The Johnson-Cook constitutive relation is expressed as:

$$\sigma = (\mathrm{A} + B\varepsilon^n)(1 + C\ln\dot{\varepsilon})\left(1 - \left|\frac{T-T_r}{T_m-T_r}\right|^m\right) \quad (1)$$

Where σ is the equivalent flow stress; A, B, C, n and m are material constants; ε is the equivalent plastic strain; $\dot{\varepsilon}$ is the equivalent plastic strain rate; T is the current temperature; T_m is the melting temperature; T_r is the reference temperature [3, 6].

Due to its ability to realistically describe the behavior of metallic materials under dynamic loading, this model is one of the most commonly used in the analysis of explosive processes and high-speed impacts [3, 4, 6, 7].

To describe the behavior of the explosive charge, the Jones–Wilkins–Lee (JWL) equation of state was applied. The JWL equation relates the pressure of the detonation products to their relative volume and specific internal energy and can be expressed as:

$$P = A\left|1 - \frac{\omega}{R_1 V}\right| e^{-R_1 V} + B\left|1 - \frac{\omega}{R_2 V}\right| e^{-R_2 V} + \frac{\omega E_0}{V} \quad (2)$$

Where P is the pressure of the detonation products; V is the relative volume; A, B, R_1, R_2 и ω are material constants; E_0 is the specific internal energy in explosiv [3, 6].

Due to its reliability in modeling detonation processes, the JWL equation represents a standard in numerical analyzes of explosive charges [6, 7].

The parameters of the used models were taken from the available literature and appropriate databases, with the values selected for the materials used in this research.

2.2 Experimental setup - Static penetration test

Experimental tests (static penetration tests in real conditions) were conducted with the aim of determining the penetration characteristics of an explosively formed projectile and validating the results obtained by numerical simulation. The experimental setting was defined so that the warhead was placed at a standoff distance of 1,5 m in relation to the armored target.

A schematic representation of the experimental setup for penetration tests is shown in Figure 2, and the materials of the target and warhead components are listed in Table 2.

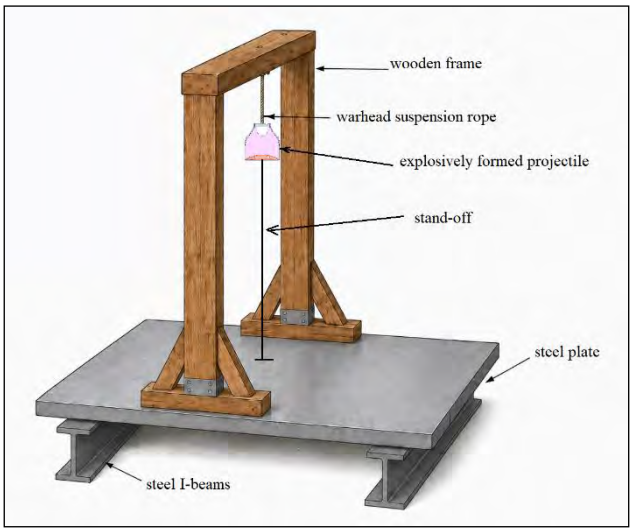


Figure 2. Setup for static penetration test

For the purposes of numerical simulation, the chosen explosive is octogen (HMX), while for the needs of real explosive charging, this explosive was used with the addition of phlegmatizer in a small mass fraction (an inert component that enables safe pressing of HMX). The presence of the inert component was neglected for the purposes of the numerical simulation.

Experimental tests included static penetration tests with two different target configurations. In configuration I, penetration through a 50 mm thick steel plate was analyzed, while in configuration II, penetration through a 100 mm thick steel plate was tested. The change in plate thickness was performed in order to see the full penetration effect, i.e. to estimate the depth of penetration. In this way, it was possible to analyze the behavior of an explosively formed projectile upon impact with a homogeneous obstacle. The obtained experimental results represent the basis for assessing the accuracy of the numerical model and analyzing the degree of agreement between the numerical and experimental approaches.

Table 3. Parameters for static penetration test

Parameter	
Stand-off	1,5 m
Configuration I	Armor steel, thickness 50 mm
Configuration II	Armor steel, thickness 100 mm
Result	Penetration depth - 85 mm

3. RESULTS AND DISCUSION

3.1. Process of explosivly formated projectile

The numerical model was developred asa a 2D axisymmetric model, with one quarter of the model considered due to the axis symmetry of the problem. A Lagragian formulation was applied to both the liner and the target. The liner was discretized with 201 elements, while the target mesh consisted of 400 elements. Numerical simulation was then used to analyze the process of forming an explosively formed projectile after the initiation of the explosive charge. During detonation, intense plastic deformation of the copper disk occurs, forming a compact projectile capable of maintaining stability during flight and achieving significant penetration when hitting a target.

Folowing the initiation of the explosive charge, the pressure of the detonation produts causes deformation and acceleration of the liner, leading to EFP formation. The stage of projectile formation at different time instants are shown in Figure 3. Figure 4 shows a ilustration of the formed penetrator.

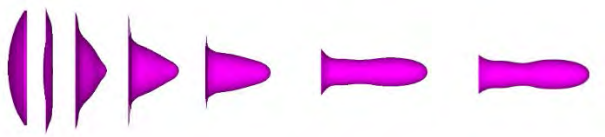


Figure 3. Penetrator formed over numerical simulation

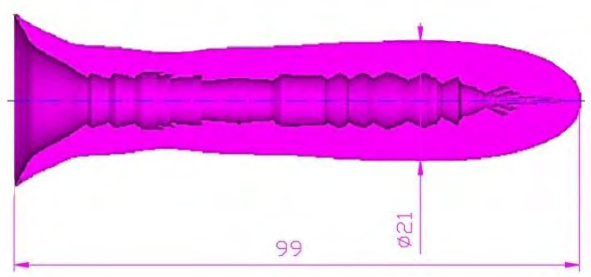


Figure 4. Penetrator formed over numerical simulation

Based on the results of the numerical simulation, the basic geometric and ballistic characteristics of the formed projectile were determined, including its shape, characteristic dimensions, mass and speed immediately before hitting the target. These characteristics represent the basis for the analysis of the penetration process and comparison with the test results in real conditions.

Table 4. Caharacteristics of EFP from numerical simulation

Parameter	Value
Projectile length	99 mm
Projectile diameter	21 mm
Projectile mass	295 g
Projectile speed	2050 m/s
Penetration depth	90 mm

3.2. Penetration of the EFP

Penetration characteristics of the explosively formed projectile were analyzed by numerical simulation and experimental through static penetration tests.

The results of the numerical simulation showed that an explosively forming projectile of a given geometry achieves penetration through a steel target (homogeneous armor plate) of approximately 90 mm, which was also confirmed in the real conditions of the penetration test (of about 85 mm).

By comparing the results of numerical simulation and experimental tests, a very good agreement can be observed regarding the achieved penetration depth. The obtained agreement between the numerical and experimental results confirms that the applied numerical model can reliably predict the penetrating characteristics of the explosively formed projectile under the analyzed test conditions. Minor deviations that may occur between numerical and experimental results can be attributed to the idealization of the numerical model, the limitations of the applied material models and experimental uncertainties, which is in accordance with the observations of other authors who were involved in the validation of numerical models of explosively formed projectiles [4, 5, 10].

4. CONCLUSIONS

In this paper, a comparative analysis of the results obtained by numerical simulation and experimental testing of an explosively formed projectile was performed. The analysis included the formation of the projectile, the determination of its basic geometric and ballistic characteristics, as well as the comparison of the achieved penetration depth in numerical simulation 90 mm through real static penetration tests.

The results of the numerical simulation showed that the applied model successfully describes the formation process of an explosively formed projectile and enables reliable determination of its basic characteristics. A comparative analysis with the results during real static penetration tests showed a high degree of agreement, whereby a penetration depth of about 85 mm was achieved. With a 100 mm thick steel plate, the projectile did not achieve complete penetration, while with a configuration with a single 50 mm thick steel plate, it completely penetrated the plate and continued to penetrate the soil base, with the total depth of penetration remaining practically unchanged.

The scientific contribution of this work is reflected in the validation of the numerical model through real static penetration tests. The confirmed agreement between numerical simulation and static penetration tests indicates that numerical simulations can be successfully used for the analysis and optimization of the construction of explosively formed projectiles, thereby reducing the need for a large number of complex and expensive polygon tests in the initial stages of their development.

The results presented in this paper represent the basis for further research aimed at the optimization of the disk geometry, the choice of materials and the improvement of numerical models, with the aim of increasing the efficiency and reliability of explosively formed projectiles in different application conditions.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] ESER GÜREL: Modeling and simulation of shaped charges, Middle East Technical University, 2009. [2] W. P. WALTERS, J.A. ZUKAS: Fundamentals of shaped charges, New Zork, 1943.

[2] WEICKERT, C. A., GALLAGHER, P. J.: "Penetration of Explosively Formed Projectiles," International Journal of Impact Engineering, Vol. 14, 1993.

[3] WU, J., LIU, J., DU, Y.: "Experimental and Numerical Study on the Flight and Penetration Properties of Explosively-Formed Projectile," International Journal of Impact Engineering, Vol. 34, No. 7, 2007.

[4] CARDOSO, D., TEIXEIRA-DIAS, F.: "Modelling the Formation of Explosively Formed Projectiles (EFP)," International Journal of Impact Engineering, Vol. 93, 2016.

[5] HUSSAIN, G. et al.: "Analytical Performance Study of Explosively Formed Projectiles," Journal of Applied Mechanics and Technical Physics, Vol. 54, 2013.

[6] YANG, D., LIN, J.: "Numerical Investigation on the Formation and Penetration Behavior of Explosively Formed Projectile with Variable Thickness Liner," Symmetry, Vol. 13, No. 8, 2021.

[7] LIU, Y. et al.: "The EFP Formation and Penetration Capability of Double-Layer Shaped Charge with Wave Shaper," Materials, Vol. 13, No. 20, 2020.

[8] BORKOWSKI, J. et al.: "Application of Sintered Liners for Explosively Formed Projectile Charges," International Journal of Impact Engineering, Vol. 118, 2018.

[9] HU, F. et al.: "Impact Performance of Explosively Formed Projectile (EFP) into Concrete Targets," International Journal of Impact Engineering, Vol. 109, 2017.

[10] LI, B. Y. et al.: "Orthogonal Optimization Design and Experiments on Explosively Formed Projectiles with Fins," International Journal of Impact Engineering, Vol. 173, 2023.

[11] PYKA, D. et al.: "Experimental and Numerical Modelling of an EFP Formation and Its Interaction with the Target," Journal of Dynamic Behavior of Materials, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DETERMINING LAW COMBUSTION PARAMETERS AND THEIR PRESSURE/COMPOSITION DEPENDENCES FOR HIGH-ENERGY MATERIALS VIA NEURAL NETWORK SEQUENTIAL APPROXIMATIONS

D.A. ANUFRIEVA[1], V.S. ABRUKOV[1], VLADICA BOZIC[2], D.A. SAPYEROV[1],

I.N. Ulyanov Chuvash State University, Cheboksary, Russia, kafedra.pfn@mail.ru

[2] Poliestar, Priboj, Serbia

Abstract: *Determination of the parameters of the power law of combustion (the Saint-Robert and Vieille law, the Murali-Saint-Robert law) is the basis for the design and optimization of the composition of high-energy materials. In most works, the parameters are determined using logarithmic transformation and linear regression, which provides only averaged values. This paper proposes a new method that allows one to determine the dependence of burning rate on pressure as a function of a high degree of smoothness, and then, based on this, the dependence of the power law parameters on pressure and composition of high-energy materials. The method is based on the application of neural networks and the method of neural network sequential approximations. The method allows solving both direct and inverse problems.*

Keywords: *high-energy materials, combustion law parameters, neural network sequential approximation method.*

1. INTRODUCTION

The dependence of the burning rate of high-energy materials on pressure is described by the power law of combustion: $u = b \cdot P^n$, where u is the linear burning rate, P is pressure, b is the combustion coefficient, and n is the pressure exponent.

Traditional methods of linear regression in logarithmic coordinates yield only averaged values of b and n due to random measurement errors [1-4]

This paper proposes a method to determine the dependence u(P) as a function of a high degree of smoothness, and then the dependences n(P) and b(P).

2. RELATION BETWEEN THE COMBUSTION LAW PARAMETERS AND THE DERIVATIVE OF BURNING RATE WITH RESPECT TO PRESSURE

Let us write the combustion law of combustion for an arbitrary point i:

$$u_i = b_i \cdot P_i^{n_i} \quad (1)$$

Differentiating (1) with respect to pressure and assuming b and n are constant in an infinitesimal neighborhood of point i:

$$\left.\frac{du}{dP}\right|_{P_i} = \frac{d}{dP} \cdot \left(b_i \cdot P_i^{n_i}\right) \quad (2)$$

Since b in an infinitesimal neighborhood of point i does not depend on P, it can be taken outside the derivative sign:

$$\left.\frac{du}{dP}\right|_{P_i} = b_i \cdot \frac{d}{dP} \cdot P_i^{n_i} \quad (3)$$

or

$$\left.\frac{du}{dP}\right|_{P_i} = b_i \cdot n_i \cdot P_i^{n_i - 1} \quad (4)$$

Now we divide both sides of equation (4) by u_i, where $u_i = b_i \cdot P_i^{n_i}$

After simplification we obtain:

$$\frac{1}{u_i} \cdot \left.\frac{du}{dP}\right|_{P_i} = \frac{n_i}{P_i} \quad (5.3)$$

$$n_i = \left.\frac{du}{dP}\right|_{P_i} \cdot \frac{P_i}{u_i} \quad (6)$$

$$b_i = \frac{u_i}{P_i^{n_i}} \quad (7)$$

3. METHOD OF NEURAL NETWORK (NN) SEQUENTIAL APPROXIMATIONS

The method consists of three steps:

1. The first neural network is trained on the original noisy experimental data of burning rate versus pressure.
2. The second neural network is trained on the output of the first network and refines the initial approximation.
3. The third neural network (NN3) is trained on the output of the second network, further smoothing the function.

The Deductor platform was used to create the models. A feedforward neural network was employed: an input layer (6 neurons: pressure, percentage of ammonium perchlorate, particle size of ammonium perchlorate, percentage of PEBA binder, percentage of aluminum, percentage of lithium fluoride), one hidden layer (3–5 neurons), and an output layer (1 neuron – burning rate).

Corresponding author: I.N. Ulyanov;
E-mail: kafedra.pfn@mail.ru
DOI: 10.5937/OTEH26050A

The activation function was the hyperbolic tangent with a steepness of 0.2.

The data were randomly split: 80% for training, 20% for testing. The loss function was the mean squared error. Training was stopped when the error reached a minimum. No overfitting was observed.

4. MODELING METHODOLOGY AND RESULTS

The models were created based on experimental data for seven compositions (author: Vladica Bozic), Table 1. The compositions differed in the content of ammonium perchlorate, its particle size, PEBA binder, aluminum, and lithium fluoride. The pressure range was from 1.45 to 17.39 MPa. The burning rate was measured using the "small motor" method [5,6].

Table 1. HEM compositions used for model development.

Composition	Pressure range, MPa
AP particle size = 120 µm, AP = 78%, PEBA = 12%, Al = 10%, LiF = 0%	2.02–15.92
AP particle size = 120 µm, AP = 77%, PEBA = 12%, Al = 10%, LiF = 1%	1.45–11.17
AP = 85%, AP particle size = 90 µm, PEBA = 15%, Al = 0.5%, LiF = 0%	1.75–10.26
AP = 85%, AP particle size = 200 µm, PEBA = 15%, Al = 0.5%, LiF = 0%	2.81–17.39
AP particle size = 200 µm, AP = 84.8%, PEBA = 15%, Al = 0.5%, LiF = 0.2%	2.29–6.80
AP particle size = 200 µm, AP = 84.5%, PEBA = 15%, Al = 0%, LiF = 0.5%	2.14–9.35
AP particle size = 200 µm, AP = 84%, PEBA = 15%, Al = 0.5%, LiF = 1%	2.27–6.54

The available experimental data were organized in a special table format. A part of the organized data table is shown in Table 2.

Table 2. Example of organized data (fragment).

Pressure, Mpa	Burning rate, mm/s	AP, %	AP size, µm	PEBA, %	Al, %	LiF, %
2,02	5,32	78	120	12	10	0
2,11	5,51	78	120	12	10	0
2,22	5,73	78	120	12	10	0
3,33	5,94	78	120	12	10	0
4,45	6,16	78	120	12	10	0
5,58	6,39	78	120	12	10	0
,,,	,,,	,,,	,,,	,,,	,,,	,,,
15,38	15,08	78	120	12	10	0
15,92	15,08	78	120	12	10	0
1,45	4,1	77	120	12	10	1
1,48	4,11	77	120	12	10	1
1,51	4,12	77	120	12	10	1
1,57	4,17	77	120	12	10	1
1,66	4,25	77	120	12	10	1
1,72	4,36	77	120	12	10	1
,,,	,,,	,,,	,,,	,,,	,,,	,,,
10,22	8,71	77	120	12	10	1
10,68	8,88	77	120	12	10	1
11,17	9,05	77	120	12	10	1

The structure of the NN3 model is shown in Figure 1

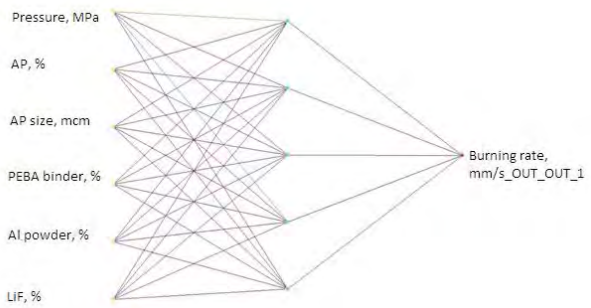


Figure 1. The structure of the NN3 model

Curves of n(P) and b(P) were constructed for all seven compositions and are depicted in Figure 2.
Figure 2 presents curves of the dependences of the power-law combustion parameters *n* and *b* on the row number in Table 2, which corresponds to seven different compositions of HEM. The ordinate axis shows the values of *n* (upper curves) and *b* (lower curves); the abscissa axis shows the row number, sequentially covering all compositions and all pressures within each composition. The blue curves show the values calculated directly from the initial experimental data u(P), while the red curves show the values obtained using the method of neural network sequential approximations (NNSA) after smoothing.

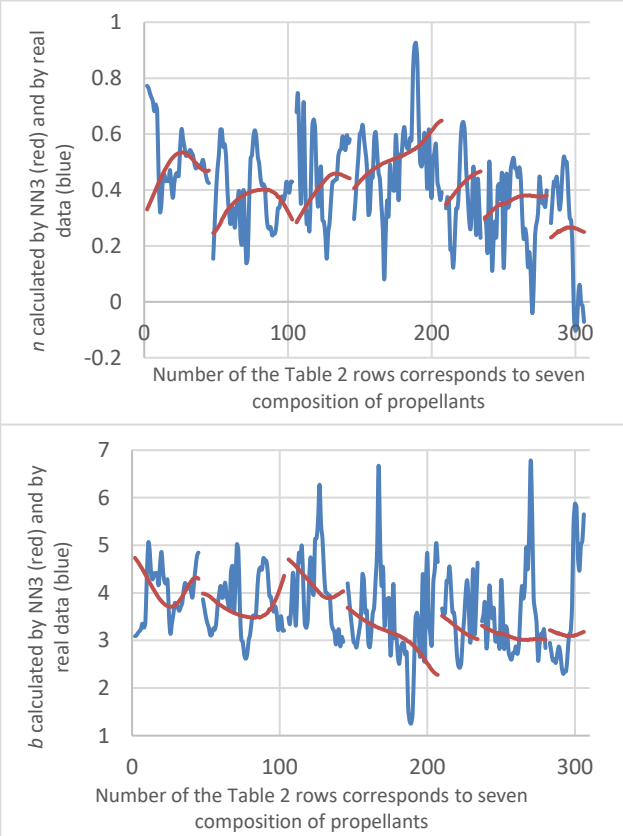


Figure 2. Curves of n(P) and b(P)

Notes:

1. Character of the blue curves (initial data).

The blue curves, obtained by direct differentiation of the noisy experimental dependences u(P), exhibit a pronounced chaotic character:

- For the exponent n, the values fluctuate over a wide range — approximately from −0.2 to 1.0 which is physically unjustified.
- For the coefficient b, the scatter ranges from 1×10−6 to 7×10−6, and the curve contains numerous sharp peaks and dips.

Such behavior is due to the fact that numerical differentiation amplifies random measurement errors, which are always present in experimental data. As a result, the blue curves are unsuitable for analysis of combustion regularities and cannot be used for solving practical problems.

2. Character of the red curves (NNSA method).

The red curves, obtained after sequential application of three neural networks, are smooth functions of a high degree of smoothness. They are free from random outliers and reflect the fundamental regularities of the combustion process.

For the exponent n:

- The values of n lie in a physically reasonable range of 0.2–0.6, which corresponds to typical values for composite solid propellants.
- A smooth variation of along the abscissa axis is observed: at the beginning (composition 1, low pressures) n≈0.35; then n increases to 0.5–0.55 in the middle of compositions 1–2; further, on passing to compositions 3–4 (with lower aluminum content and larger AP particle size), n decreases somewhat to 0.4; for the final compositions (5–7), n stabilizes in the region of 0.4–0.5.
- The transitions between compositions (at the boundaries of rows 46, 104, 144, 208, 235, 281) are smooth, without jumps, which indicates good generalizing ability of the neural network model.

For the coefficient b:

- The values of b are in the range of ~3×10−6–5×10−6.

The red curve b shows a general tendency to decrease from the beginning to the end: in the first compositions b≈4.5×10−6, and by the last compositions it decreases to 3×10−6.

As for n, the transitions between compositions are smooth, which allows analysis of the influence of composition on the coefficient b.

3. Main regularities following from Figure 2

Physical validity**.** The red curves fully agree with classical concepts of HEM combustion: 0<n<1, and b varies within a narrow range. The blue curves do not demonstrate such regularities.

Influence of composition**.** The changes in n and b correlate with composition variations:

- The introduction of the depressant LiF (compositions 2, 5, 7) leads to a certain decrease in n compared with compositions without LiF (compositions 1, 3, 4).
- An increase in aluminum content (compositions 1, 2 — 10% Al) gives higher values of n compared with compositions with low Al content (compositions 3, 4 — 0.5% Al).
- A change in the dispersity of ammonium perchlorate (90, 120, 200 μm) also affects n and b, which is reflected in smooth shifts of the red curves.

4. Identification of combustion regimes.

The smooth dependences n(P) within each composition allow identification of pressure regions where the combustion mechanism changes. For example, a change in the slope of the n curve may indicate a change in the dominant phase (gas or condensed).

Figure 2 clearly demonstrates the effectiveness of the NNSA method. The blue curves, constructed from "raw" experimental data, are unsuitable for analysis due to chaotic scatter. The red curves, obtained after neural network smoothing, represent smooth, physically consistent dependences of n and b on pressure and composition.

They allow:

- correct determination of the power-law combustion parameters;
- analysis of the influence of composition on these parameters;
- solution of inverse problems for targeted adjustment of HEM formulations.

Thus, Figure 2 confirms that the proposed method of neural network sequential approximations is a reliable tool for processing noisy experimental data and extracting hidden physical regularities from them.

Dependences of n and b on the percentage of aluminum and lithium fluoride at different pressures were obtained also. The results allow solving inverse problems: for example, determining how much lithium fluoride is needed to obtain a desired value of n at a given pressure. It will be presented later.

5. CONCLUSIONS AND PROSPECTS

The developed methodology fundamentally improves the processing of experimental data on the dependence of burning rate on pressure and the determination of the parameters of the power law of combustion and their dependence on pressure.

The analysis of n(P) and b(P) helps to identify different combustion regimes. The method of neural network que approximations provides an effective approach for eliminating random errors in experimental measurements.

The method can be applied to any experimental data contaminated with random errors, especially when the first derivative is important for process analysis.

A brief summary of the methodology presented in this paper was previously presented in [7-9].

The general principles of this methodology are presented in [10-14].

References

[1] Zeldovich, Ya. B., Barenblatt, G. I., Librovich, V. B., & Makhviladze, G. M. (1985). The Mathematical Theory of Combustion and Explosions. Consultants Bureau.

[2] Kuo, K. K. (2005). Principles of Combustion (2nd ed.). Wiley.

[3] Shen, M., & Yetter, R. A. (2017). Fundamentals of High-Energy-Density Materials. In Energetic Materials: From Cradle to Grave (pp. 1–34). Royal Society of Chemistry. https://doi.org/10.1007/978-3-319-59208-4

[4] Beckstead, M. W., Derr, R. L., & Price, C. F. (1970). A Model of Composite Solid-Propellant Combustion Based on Multiple Flames. AIAA Journal, 8(12), 2200–2207. https://doi.org/10.2514/3.6087

[5] Milos, V.M., & Bozic, V.S. (1998). Small Motor Measurement Method for Determining Burning Rate of Solid Rocket Propellants, AIAA Paper 98-3390.

[6] F. Maggi, L.T. DeLuca et al. Burn-rate Measurement on Small-scale Rocket Motors. Defence Science Journal, Vol. 56, No. 3, July 2006, pp. 353-367.

[7] Darya Anufrieva, Victor Abrukov, Vladica Bozic, Daniil Sapyorov. Determination of Combustion Law Parameters from Noisy Experimental Data Using a Novel Sequential Neural Network Approach. XVI International Symposium on "Combustion and Plasma Chemistry. Physics and Chemistry of Carbon and Nanoenergetic Materials," October 1-3, 2025, Almaty, Kazakhstan, pp. 33-36

[8] Darya Anufrieva, Victor Abrukov, Vladica Bozic, Daniil Sapyorov. Determination of Combustion Law Parameters Based on Noisy Experimental Data. Book of Abstract of 15th International Workshop on Combustion and Propulsion (IWCP) Advanced Design, Additive Manufacturing, and Characterization of Modern Energetic Materials Advanced Design, 7-9 July 2025, Pescara, Italy, p. 1

[9] Victor Abrukov and Darya Anufrieva, Determination of Combustion Law Parameters Based on Noisy Experimental Data. October 2025, DOI: 10.13140/RG.2.2.12141.04328, Conference: International Conference on Numerical Combustion, Rome, Italy – 14-17 October, 2025

[10] Abrukov V, Pang W, Anufrieva D. Methodology for Studying Combustion of Solid Rocket Propellants using Artificial NeuralNetworks. Ann Adv Chem. 2024; 8:001-007

[11] Abrukov VS, Pang W, Anufrieva DA. Neural networks are a methodological basis of materials genome. Trends Comput Sci InfTechnol. 2023; 8(1): 012-015. DOI: https://dx.doi.org/10.17352/tcsit.000063.

[12] Pang, W.; Abrukov, V.; Anufrieva, D.; Chen, D. Burning Rate Prediction of Solid Rocket Propellant (SRP) with High-Energy Materials Genome (HEMG). Crystals 2023, 13, pp. 237-252. https://doi.org/10.3390/cryst13020237.

[13]. Doherty R.M., Abrukov V.S., Boukouvalas Z., Baek S., Terreux R. Application of AI and ML to Novel Energetic Materials // Energetic Materials and Techniques: Advances in Chemical Propulsion and Power Generation : in 2 vol. / ed. by L.T. DeLuca, R.M. Doherty. — Wiley, 2026. — Chapter 1. — ISBN 978-3-527-35545-7. — https://doi.org/10.1002/9783527853267.ch01

[14]. Abrukov V., Pang W., Anufrieva D., Sapyorov D. Neural Networks in the Development of High-Energy Materials Genome // Energetic Materials and Techniques: Advances in Chemical Propulsion and Power Generation : in 2 vol. / ed. by L.T. DeLuca, R.M. Doherty. — Wiley, 2026. — Chapter 3. — ISBN 978-3-527-35545-7. — https://doi.org/10.1002/9783527853267.ch03

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF GROOVE GEOMETRY ON THE INTERNAL BALLISTICS OF A MONOLITHIC .308 CALIBER BULLET

IGOR BABIĆ
Prvi Partizan, Užice, igor.babic@prvipartizan.com

MIHAILO ERČEVIĆ
Prvi Partizan, Užice, mihailo.ercevic@prvipartizan.com

Abstract: *With the growing awareness of environmental protection and the harmful effects of lead, regulatory changes have been introduced, accompanied by a clear trend toward the gradual elimination of conventional "Soft Point" and "FMJ" bullets and their replacement with monolithic ("solid") bullets made of copper alloys. These bullets are characterized by a reduced negative impact on human health and the environment, as well as by certain properties favorable for sporting applications. A drawback of the monolithic design is the absence of a softer (lead) core, which results in increased resistance during the engraving of the bullet into the barrel rifling. To mitigate this issue, grooves are formed on the bullet surface, whose function is to accommodate and redistribute displaced material during the engraving process into the rifling grooves. This paper analyzes the influence of the geometry of these grooves on the internal ballistics of a .308 caliber "Wadcutter Match" bullet, with the aim of establishing a basis for further development and optimization of monolithic bullet design.*

Keywords*: bullet, solid , rifling, grooves, internal balistics.*

1. INTRODUCTION

Monolithic projectiles have become increasingly attractive alternatives to conventional lead-core bullets due to their environmental compatibility, high structural integrity, and superior terminal performance. However, their larger bearing surface and increased engraving resistance often produce higher chamber pressures than conventional projectiles, making the optimization of projectile geometry an important aspect of internal ballistic design.

This study experimentally investigates the influence of material displacement groove geometry on the internal ballistic performance of monolithic bullets. Three groove geometries with different engraved material volumes were manufactured while maintaining identical projectile mass, calibre, and external dimensions. The projectiles were tested using two propellant types under controlled laboratory conditions, and chamber pressure and muzzle velocity were recorded.

The experimental results demonstrate that groove geometry significantly affects peak chamber pressure while producing only minor changes in muzzle velocity. The projectile with the smallest engraved material volume exhibited the lowest chamber pressure and the highest Internal Ballistics Efficiency Index, indicating the most favourable balance between ballistic performance and mechanical loading of the firearm. Additional experiments investigating the influence of groove number shows that the pressure/velocity relation curve peaks at a optimal point and any additional decrease in engraving resistance does not improve end results.

The results suggest that engraved material volume represents a practical geometric design parameter for optimizing monolithic projectiles. The proposed Internal Ballistics Efficiency Index provides an additional engineering criterion for comparing projectile designs and may support the future development of pressure-optimized monolithic ammunition.

2. MATERIAL DISPLACEMENT GROOVE GEOMETRY

In the first stage of the experimental investigation, three groove geometries were selected to evaluate the influence of groove profile on the internal ballistic performance of monolithic projectiles. All projectile variants were manufactured with identical external dimensions, mass, material, and manufacturing tolerances. The only variable was the geometry of the grooves.
The investigated groove configurations were:

Type A – Raised Bands: The bearing bands have a diameter equal to the nominal projectile diameter, while the recessed regions between the bands have a reduced diameter (Figure 1).
Type B – Depressed Grooves: Conventional grooves are machined directly into the projectile body, resulting in a continuous bearing surface interrupted by circumferential grooves (Figure 2).
Type C – Wavy Profile: The bearing surface follows a continuous wave-shaped profile, providing a gradual transition between the bearing bands and recessed regions (Figure 3).

Corresponding author: I. Babić;
E-mail: igor.babic@prvipartizan.com
DOI: 10.5937/OTEH26160B

The objective of this comparison was to determine how the groove profile influences projectile engraving resistance, chamber pressure, and muzzle velocity while maintaining all other projectile parameters constant.
Following the identification of the most favourable groove geometry, a second series of experiments was conducted to investigate the influence of the **number of** grooves on internal ballistic performance.

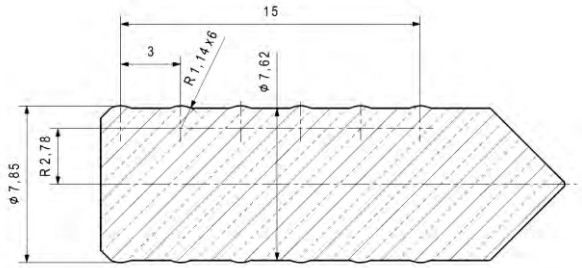


Figure 1. Groove design. A.

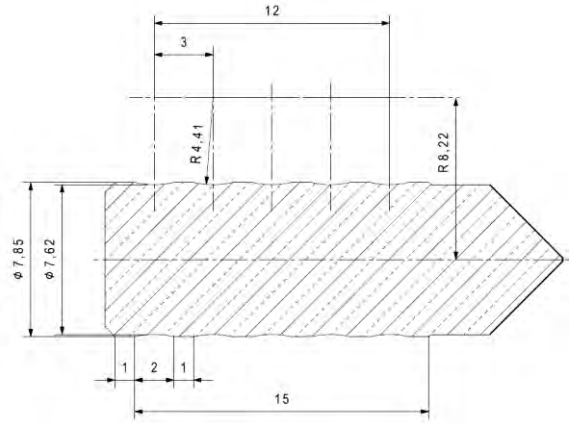


Figure 2. Groove design. B.

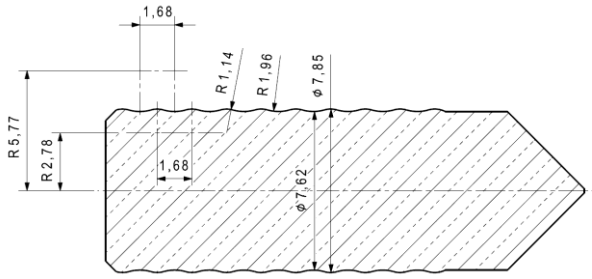


Figure 3. Groove design. C.

3. EXPERIMENT SETUP

The experimental investigation was conducted at the Prvi Partizan Užice company inside a controlled indoor ballistic testing facility. All firing tests were performed under identical laboratory conditions in order to minimize the influence of external factors on the measured ballistic parameters.

The test ammunition consisted of .308 cartridges loaded with CNC lathe turned monolithic brass projectiles. All projectiles were manufactured from CuZn39Pb2 alloy, maintained identical external dimensions, and were manufactured using identical machining procedures. The investigated variables were limited exclusively to the geometry and number of grooves, allowing the influence of these parameters on internal ballistic performance to be isolated.

The cartridges were assembled using identical cartridge cases, primers, and carefully controlled propellant charges. Two propellant types, **SB511** and **SB520**, were selected to investigate whether the observed influence of projectile geometry remained consistent under different combustion characteristics.

All projectiles were fired from the same test barrel mounted in a rigid fixture. Using a single barrel eliminated variations caused by manufacturing tolerances, barrel wear, chamber dimensions, and rifling characteristics, ensuring that differences in measured chamber pressure and muzzle velocity resulted solely from changes in projectile geometry.

Peak chamber pressure was measured using a piezoelectric pressure transducer installed in accordance with the C.I.P standard pressure measurement procedure for small-calibre ammunition.

Projectile muzzle velocity was measured using a chronograph.

To evaluate experimental repeatability, 5 shots were fired for each projectile configuration and propellant combination. Mean values, standard deviations, and coefficients of variation were subsequently calculated for both chamber pressure and muzzle velocity. The use of repeated measurements enabled the assessment of measurement uncertainty while reducing the influence of random experimental variation.

The experimental procedure was performed in two successive stages. The first stage investigated the influence of three different groove geometries while maintaining identical projectile mass. Following identification of the most favourable groove configuration, the second stage examined the influence of the number of grooves and projectile mass reduction through the introduction of an internal cavity. This sequential approach enabled the individual effects of groove geometry, groove number, and projectile mass on internal ballistic performance to be evaluated independently.

Powder was manufactured by Milan Blagojević Namenska factory in Lučani, all powder is used from the same lot to exclude manufacturing differences.
Overall bullet length is $60{,}5^{-0{,}5}$ mm for all tests.

Table 1. Powder.

Rb.	Powder:	Lot:
1.	SB511	Ser. 25339
2.	SB520	Ser. 25155

Table 2. Measuring equipment

Rb.	Equipment	Manufacturer/model
1.	Chronograph	KISTLER
2.	Pressure sensor	KISTLER 6215
3.	Vernier caliper	HOREX 150mm
4.	Scale	RCBS Charge master 1500

4. EXPERIMENTAL RESULTS

The internal ballistic performance of the investigated monolithic projectiles was evaluated by measuring the maximum chamber pressure and muzzle velocity for each projectile configuration.

The experimental investigation was performed in two stages. In the first stage, three groove geometries (Types A, B, and C) having identical mass and external dimensions were compared in order to isolate the influence of groove profile on internal ballistic performance. Following the identification of the most favourable groove geometry, the second stage investigated the influence of the number of grooves and projectile mass reduction.

For each projectile configuration, multiple firing tests were conducted to ensure measurement repeatability. The recorded chamber pressures and muzzle velocities were subsequently processed to obtain average values and standard deviations, which were used for the comparative analysis presented in the following sections.

The experimental results are first discussed in terms of the influence of groove geometry on chamber pressure and muzzle velocity. Subsequently, the effects of groove number and projectile mass are analysed using the additional experiments performed with the two propellant types. Finally, the Internal Ballistics Efficiency Index (IBEI) is introduced as a supplementary engineering parameter for evaluating the relationship between muzzle velocity and peak chamber pressure.

4.1 Influence of Groove Geometry

For the first experiment iteration a load of SB511 powder and charge of 2,75g was used to record obtained pressure and adjust the powder charges according to maximum pressure.

Table 3. Results of first experiment iteration.

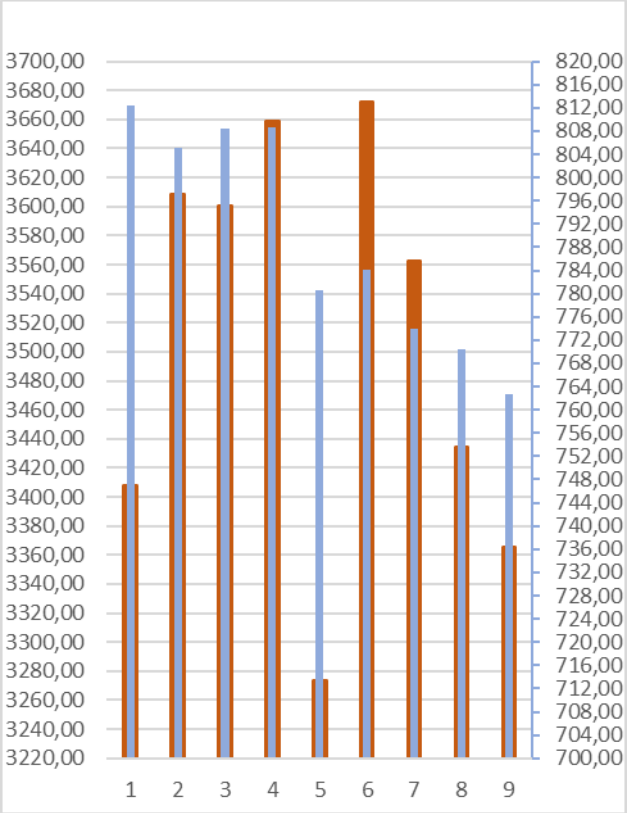
Rb.	G.Geometry	OAL [mm]	Velocity [m/s]	Pressure [bar]
1.	A	60,48	814	3770
2.	B	60,46	804	4040
3.	C	60,47	818	4199

For the second experiment iteration the powder charge was adjusted for expected max pressure of ~3600 bar as a safe upper limit, for each groove geometry and both SB511 and SB520 powders.

Table 4. Results of second experiment iteration

Rb	G. Geometry	Powder	Powder charge [gram]	Velocity [m/s]	Pressure [bar]
1.	A	SB520	3	812,32	3408,00
2.	B	SB520	3	805,04	3608,86
3.	C	SB520	3	808,31	3600,37
4.	A	SB511	2,75	808,72	3658,69
5.	A	SB511	2,65	780,56	3273,32
6.	B	SB511	2,65	784,10	3671,97
7.	B	SB511	2,60	773,98	3562,72
8.	C	SB511	2,60	770,51	3434,11
9.	C	SB511	2,55	762,64	3365,31

The results of the second experiment iteration is a correlation between maximum pressure and obtained velocity for each powder charge and grove geometry type combination used.



Graph 1. Relationship between velocity and pressure.
Orange bar, left side – Pressure [bar]
Blue bar, right side – Velocity.[m/s]

From the experiment results and accompanying graph it can be concluded that the groove geometry model "A" provides the highest velocity compared to max pressure meaning that a bullet with "A" groove geometry can be propelled at higher powder loads to achieve higher velocity while still being under the maximum safe chamber pressure.

For the SB520 powder from the graph can be concluded that under the same load od 3g of powder for all three groove geometries used, "A" type achieves the highest velocity for a significant difference in pressure compared to "B" and "C" types. The measured velocity is about the same for all three, but the pressure is the most impacted by groove geometry.

The experiment also shows us that for powder SB511 and "A" g. geometry a difference in 0,1 gram of powder can drastically change the pressure-velocity relation.

Grove geometry type "B" shows us a unfavorable correlation between pressure and velocity compared to "A" and "C", achieving the highest pressures with moderate velocity. Comparing 5-6 with same load "B" achieves negligible higher velocity (4m/s) than "A" but for a 400 bar higher pressure.

Groove geometry type "C" achieves marginally better results than "B", comparing 7-8 that have the same load, "C" achieves negligible lower velocity (3m/s) with ~130 bar lower pressure.

4.2Analysis of first part

From an internal ballistic point of view the geometry which achieves the highest velocity and lowest pressure is the most favorable. Increasing bullets starting velocity achieves flatter trajectory and greater range, while keeping safe operating pressures does not compromise firearm integrity and safety. Here we can conclude that within the investigated geometries, Type A exhibited the most favorable pressure–velocity relationship.

While all three designs have the same material, overall length, weight, diameter and tolerances, the only difference being the geometry of the grooves.

The grooves here have two major functions:

1. Gas sealing
2. Engraving the bullet into barrel rifling and spinning the bullet.

Increasing the sealing of gases has the benefit of better efficiency where more energy of burning powder goes directly into propelling the bullet forward increasing its velocity, but preventing gas leak past the projectile could cause a increase in peak pressure.

Increasing how much material engages the rifling and is directly engraved into it causes the increase in the frictional resistance of the projectile which decreases its velocity while increasing the peak pressure.

For analysis the nominal barrel diameters are used according to C.I.P. Barrel tolerance and wear are not considered although in real world they affect the depth of rifling grooves and directly affect the engraving resistance.

The volume of "engraved" material for all three groove geometry types is :
A - 11,322 mm^3
B - 26,623mm^3 velocit
C- 21,908mm^3

We can see that type "A" has half the volume of the projectile engraving into rifling compared to other two types. Comparing them for minimal engraved volume we have:

1. A - the lowest,
2. C - second
3. B - the highest

Comparing this to the results we have a clear trend of "A" being the best, and "C" being only marginally better than "B". These observations suggest a positive relationship between engraved material and the projectiles resistance to maximum pressure and velocity.

The results suggest that engraved volume may be treated as an independent design variable for predicting the internal ballistic behavior of monolithic projectiles. Reducing the volume of material displaced during rifling engraving decreases peak chamber pressure while increasing muzzle velocity, making engraved volume a promising optimization parameter for future projectile development.

4.3 Influence of the Number of Grooves

Following the identification of the most favorable groove geometry, a second series of experiments was conducted to investigate the influence of the number of grooves on the internal ballistic performance of monolithic projectiles.

The "A" type groove geometry with raised bands was chosen as the optimal of analyzed variants.
A model with decreased mass and same number of grooves was also included in testing to establish a relation for future experiments.

There were 5 new projectiles for this experiment:

1. A - 6 bands on the bullet, 8g mass (as used before)
2. A.1 - 3 bands evenly spread on the bullet, ~8g mass.
3. A.2 - 3 bands, two at the back and one on the front of the bullet, ~8g mass.
4. A.3 - 4 bands evenly spread on the bullet, ~8g mass.
5. A.4 - 5 bands evenly spread on the bullet, ~8g mass.
6. A.5 - 6 bands evenly spread on the bullet, with lightening hole, ~7,2g mass.

The average values of maximum chamber pressure and muzzle velocity obtained for each projectile configuration are presented in Tables 5 and 6.

Table 5. Second experiment 1. Iteration, powder SB520.

Rb	G. Geometry	Number of bands	Powder charge [gram]	Velocity [m/s]	Pressure [bar]
1.	A	6	3	809,05	3433,67
2.	A.1	3 even	3	807,75	3271,33
3.	A.2	3	3	809,69	3339,67
4.	A.3	4	3	806,90	3326,33
5.	A.4	5	3	812,41	3402,67
6.	A.5	6+hole	3	818,32	3093,33

Table 6. Second experiment 2. Iteration, powder SB511.

Rb	G. Geometry	Number of bands	Powder charge [gram]	Velocity [m/s]	Pressure [bar]
1.	A	6	2,65	775,41	3177,00
2.	A.1	3 even	2,65	771,98	2972,33
3.	A.2	3	2,65	765,51	2911,00
4.	A.3	4	2,65	735,67	2427,67
5.	A.4	5	2,65	773,64	3022,33
6.	A.5	6+hole	2,65	779,42	2693,00

During the initial stage of projectile motion, part of the projectile material is plastically displaced into the rifling lands. The mechanical work required for this deformation increases the resistance opposing projectile motion and contributes directly to the generation of peak chamber pressure. Consequently, projectile designs requiring a smaller volume of material to be displaced during engraving exhibit lower chamber pressures while maintaining comparable muzzle velocities.

The experimental results demonstrate that decreasing the number of grooves (and the engraved material volume at the same time) does not produce a proportional reduction in chamber pressure. Although the engraving resistance is decreased measured pressure differences between the investigated groove configurations remain relatively modest. This observation indicates that there is a limit at a certain point where decreasing the engraving resistance stops affecting pressure/velocity relation at an favourable way.

This trend was consistently observed for both propellants. Although the absolute chamber pressures and muzzle velocities differed due to the different combustion characteristics of SB511 and SB520 powders, the relative influence of projectile geometry remained similar. The consistency of these observations under two independent propellant conditions supports the conclusion that the results are primarily due to bullet geometry rather than propellant-specific.

The projectile incorporating an internal cavity exhibited an additional reduction in chamber pressure while simultaneously producing the highest muzzle velocity among the investigated configurations. This behaviour can be attributed primarily to the reduction in projectile mass, which decreases the force required to accelerate the projectile during the initial phase of internal ballistics. The reduced inertial resistance allows the projectile to begin moving earlier, thereby increasing the available expansion volume for the combustion gases and limiting the maximum chamber pressure.

Overall, the second experimental series confirms that decreasing engraved material volume past a certain „lower limit“ does not necessarily improve internal ballistic performance. Rather, the results indicate that optimization should focus on reducing engraving resistance through careful control of engraved material volume and projectile mass while maintaining sufficient bearing surface to ensure reliable guidance and sealing within the barrel.

5.INTERNAL BALLISTIC EFFICIENCY INDEX

To evaluate the ballistic performance of the tested projectile designs from an engineering perspective, an additional parameter, referred to as the Internal balilistics efficiency index (IBEI), was introduced. While muzzle velocity and maximum chamber pressure are commonly analyzed separately, neither parameter alone fully describes the effectiveness of a projectile design. A projectile that achieves a high muzzle velocity at the expense of substantially increased chamber pressure cannot be considered equally efficient as one producing similar velocity at lower pressure.

The IBEI is therefore defined as:

$$IBEI = V/P \quad (1)$$

where:

- (V) is the average muzzle velocity (m/s),
- (P) is the average maximum chamber pressure (bar).

The IBEI represents the amount of muzzle velocity generated per unit of chamber pressure. Higher values indicate that a projectile converts the available propellant energy into muzzle velocity more efficiently while imposing a lower mechanical load on the firearm.

Table 7. IBEI for first experiment.

No.	Geometry	Powder	Charge (g)	Velocity (m/s)	Pressure (bar)	IBEI ($m \cdot s^{-1}$/bar)
1	A	SB520	3.00	812.32	3408.00	0.2384
2	B	SB520	3.00	805.04	3608.86	0.2231
3	C	SB520	3.00	808.31	3600.37	0.2245
4	A	SB511	2.75	808.72	3658.69	0.2210
5	A	SB511	2.65	780.56	3273.32	0.2385
6	B	SB511	2.65	784.10	3671.97	0.2135
7	B	SB511	2.60	773.98	3562.72	0.2172
8	C	SB511	2.60	770.51	3434.11	0.2244
9	C	SB511	2.55	762.64	3365.31	0.2266

The calculated IBEI values show us that groove geometry type A has the most favorable pressure and velocity relation of tested designs and loads, acros booth powders.

Using the experimental results obtained with the SB520 propellant, the calculated IBEI values are presented in Table 8.

Table 8. Internal ballistics efficiency index for SB520.

Projectile configuration	Mean velocity [m/s]	Mean pressure [bar]	IBEI [$m \cdot s^{-1}$/bar]
3 grooves (even)	807.75	3271.3	0.2466
3 grooves (2 rear)	809.69	3339.7	0.2425
4 grooves	806.90	3326.3	0.2426
5 grooves	812.41	3402.7	0.2387
6 grooves	809.05	3433.7	0.2356
6 grooves + cavity	818.32	3093.3	0.2646

The results demonstrate measurable differences in

pressure efficiency among the investigated projectile geometries. Although the muzzle velocities differ by less than approximately 2%, the corresponding chamber pressures vary considerably, resulting in noticeable differences in the Pressure Efficiency Index.

The projectile incorporating six grooves and an internal cavity achieved the highest PEI value (0.2646 $m \cdot s^{-1}$/bar), indicating the most favorable balance between muzzle velocity and chamber pressure. Compared with the conventional six-groove configuration, this represents an improvement of approximately 12%, primarily due to the reduction in projectile mass and the corresponding decrease in chamber pressure while simultaneously achieving a higher muzzle velocity.

Among the projectiles of identical mass, the differences in IBEI are smaller but remain significant. Configurations with reduced engraving resistance generally exhibit higher pressure efficiency, suggesting that projectile geometry influences internal ballistic performance primarily through its effect on engraving resistance rather than through changes in muzzle velocity alone.

The calculated IBEI values for the SB511 propellant are presented in Table 9.

Table 9. Internal ballistics efficiency index for SB511.

Projectile configuration	Mean velocity (m/s)	Mean pressure (bar)	PEI ($\times 10^{-3}$ $m \cdot s^{-1}$/bar)
3 grooves (even)	771.98	2972.3	0.2597
3 grooves (2 rear)	765.51	2911.0	0.2630
4 grooves	735.67	2427.7	0.3030
5 grooves	773.64	3022.3	0.2560
6 grooves	775.41	3177.0	0.2441
6 grooves + cavity	779.42	2693.0	0.2894

The calculated IBEI values exhibit larger variation than those obtained with the SB520 propellant. Although the muzzle velocities differ by approximately 5%, the corresponding chamber pressures differ by more than 700 bar, resulting in substantial differences in pressure efficiency.

Among the tested configurations, the four-groove projectile produced the highest IBEI value (0.3030$m \cdot s^{-1}$/bar). This result is primarily a consequence of its considerably lower chamber pressure while maintaining a relatively high muzzle velocity. Interestingly, this same configuration exhibited the lowest absolute velocity of all projectiles tested with SB511. Consequently, the high IBEI value should not be interpreted as evidence of superior overall ballistic performance but rather as an indication of exceptionally low chamber loading relative to the achieved velocity.

The six-groove projectile with an internal cavity also demonstrated a comparatively high IBEI (0.2894 $m \cdot s^{-1}$/bar), confirming that reducing projectile mass can improve pressure efficiency by lowering the resistance to acceleration and reducing peak chamber pressure.

The remaining projectile configurations produced IBEI values ranging from 244.1 to 263.0 $\times 10^{-3}$ $m \cdot s^{-1}$/bar, indicating relatively similar pressure efficiencies despite differences in groove arrangement. The conventional six-groove projectile exhibited the lowest IBEI, primarily because it generated the highest chamber pressure while providing only a marginal increase in muzzle velocity compared with the remaining configurations.

Overall, the SB511 results indicate that pressure efficiency is influenced by both projectile geometry and projectile mass. However, unlike the SB520 propellant, the relationship is less uniform due to the unexpectedly low chamber pressure recorded for the four-groove configuration. This observation suggests that additional mechanisms, such as engraving sequence, bearing surface interaction, or gas sealing characteristics, may contribute to the internal ballistic behavior of this particular design.

These findings indicate that the IBEI provides a useful supplementary parameter for comparing monolithic projectile designs. Unlike velocity or chamber pressure considered independently, the IBEI simultaneously accounts for both ballistic performance and mechanical loading on the firearm. Consequently, it enables a more comprehensive assessment of projectile efficiency and may serve as a practical design criterion during the optimization of future monolithic projectiles. Further investigation with a larger experimental sample is recommended to determine whether this result represents a reproducible physical phenomenon or an isolated experimental effect.

6.CONCLUSION

The results demonstrated that modifications to the groove geometry significantly influence peak chamber pressure while also producing changes in muzzle velocity. Among the three groove profiles investigated in the first experimental series, **Type A** consistently exhibited the lowest chamber pressure while maintaining the highest muzzle velocity, indicating the most favourable balance between firearm loading and ballistic performance within the investigated configurations.

Analysis of the groove geometries showed that the observed differences in chamber pressure are primarily associated with the engraved material volume rather than with the groove profile itself. Projectile configurations requiring a smaller volume of material to be plastically displaced during rifling engagement generated lower chamber pressures, indicating that engraving resistance is one of the dominant mechanical factors governing the initial stage of internal ballistics.

The second experimental series demonstrated that simply increasing the number of pressure-relief grooves does not necessarily result in proportional reductions in chamber pressure. Instead, the experimental results indicate that the effectiveness of pressure-relief grooves depends on their influence on engraved material volume and the resulting resistance to projectile engraving. Furthermore, the projectile incorporating an internal cavity exhibited lower chamber pressure and increased muzzle velocity,

confirming that projectile mass also contributes to internal ballistic performance by reducing inertial resistance during projectile acceleration.

An important finding of this study is that projectile geometry produced substantially larger relative changes in chamber pressure than in muzzle velocity. This observation demonstrates that chamber pressure can be reduced through geometric optimization without significantly compromising ballistic performance, providing valuable guidance for the future design of monolithic projectiles.

Overall, the experimental results suggest that engraved material volume represents a practical engineering parameter for the optimization of pressure-relief groove geometry. Rather than considering groove number or groove shape independently, future projectile design should focus on minimizing the volume of material displaced during rifling engraving while maintaining sufficient bearing surface to ensure reliable guidance, structural integrity, and gas sealing.

Future research should investigate a wider range of projectile geometries, calibres, materials, and propellant types, as well as develop predictive numerical and statistical models capable of estimating chamber pressure from measurable geometric parameters.

References

[1] Group of Authors., *Ekološka municija*, Prvi Partizan, Užice, 2012.

[2] Stanimir R. Arsenijević, *Hemija, opšta i neorganska, Naučna knjiga Beograd*, 1990.

[3] Tests of monolithic sniper ammunition in PPU.

[4] C.I.P. 308 Win. TAB. I Date 84-06-14 Country of Origin: US Revision 08-09-23,

[5] Pihtili, H., Yildirim, R.: Investigation of the Effect of Various Types and Features of Grooved Barrels on the Range of Firearms, Defence Science Journal, Vol. 73, No. 1, January 2023, pp. 11-19, DOI : 10.14429/

[6] Engineering design handbook ballistics series, internal ballistics of guns, headquarters, u.s.army materiel command, february 1965.

[7] Research and development of materiel, engineering design handbook, elements of armament engineering part two ballistics, headquarters, u.s.army materiel command, 1963

[8] Technical report arlc3 tr 85008, wear of projectile rotating bands II, Robert S. Montgomery, march 1985

[9] Sidra I. S. Paul, W.: Effect of rifling grooves on the performance of small-caliber ammunition, US Army Research Laboratory Aberdeen Proving Ground, MD 21005-5066

[10] Smitchko, P., J.:United States Patent,Patent No.: US 8,573,129 B1, SELF SEALING FIREARM PROJECTILE, Date of Patent: Nov. 5, 2013

[11]

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

TBX FORMULATIONS WITH ALTERNATIVE METAL FUELS, HIGH EXPLOSIVES AND OXIDIZERS

THOMAS M. KLAPÖTKE
LMU Munich, Munich, tmk@cup.uni-muenchen.de

DANICA BAJIĆ
Military Technical Institute, Belgrade, danica.bajic@mod.gov.rs

STANISŁAW CUDZIŁO
Military University of Technology, Warsaw, stanislaw.cudzilo@wat.edu.pl

MUHAMED SUĆESKA
University of Zagreb, Zagreb, muhamed.suceska@rgn.unizg.hr

Abstract: *TKX-50 is a new and very promising secondary explosive with low sensitivity (IS = 20 J, FS = 120 N) and good calculated energetic performance parameters at its theoretical maximum density (1.877 g cm^{-3} @ 298 K) such as D = 9642 m s^{-1} and pCJ = 37 GPa. Thermobaric explosives (TBX) are a combination of high explosive (RDX, HMX, TKX-50) and pyrotechnic combustion of the (usually) metallic fuel (Al, Mg, B, Be, BeH_2, H_3N-BH_3 etc.), and in some formulations, oxidizer. In this research new TBX formulations were surveyed based on: (i) conventional (RDX) and new (TKX-50) high explosives, (ii) conventional (AP) and new (TNEF) oxidizers, and (iii) conventional (Al, Mg) and new (B, Be, BeH2, H3N-BH3 etc.) metallic and non-metallic fuels.*

Keywords: *beryllium, high explosives, oxidizers, thermobarics*

1. INTRODUCTION

Thermobaric explosives (TBXs) derive their effectiveness not only from the initial detonation of the high explosive but also from the subsequent combustion of metallic fuels. Unlike conventional explosives, TBXs produce an extended pressure pulse and prolonged thermal output through post-detonation reactions, making them particularly effective in confined and semi-confined environments. The efficiency of these systems depends strongly on the ability of the metallic fuel to participate in both the oxygen-deficient (anaerobic) reactions immediately following detonation and the later oxygen-supported (aerobic) combustion that occurs as atmospheric air mixes with the expanding fireball [1-3].

Aluminum has traditionally been the preferred metallic fuel for thermobaric formulations because of its high energy density and relatively low cost. However, the practical utilization of aluminum is limited by the presence of a stable oxide layer and by the high melting and boiling temperatures of aluminum and alumina. These characteristics slow ignition and oxidation, reducing the fraction of aluminum that reacts during the relatively short lifetime of the expanding detonation products. Consequently, improving the reactivity of metallic fuels remains an important objective in the development of more efficient thermobaric explosives [4-6].

The present work summarizes two investigations which examined possible alternative strategies for improving TBX performance. The first study direction compares beryllium directly with the conventionally employed aluminum and magnesium as the sole metallic fuel in model thermobaric compositions. The second study direction investigates whether small additions of highly reactive metals - specifically beryllium or cerium - can act as activators that enhance the performance of aluminum without replacing it completely [6-10]. Together, these studies provide new insight into the role of metallic fuel chemistry during both the anaerobic and aerobic stages of thermobaric explosions.

This work also presents effects of introducing new oxidizer TNEF as a replacement for AP in thermobaric formulation.

2. METAL FUELS AND HIGH EXPLOSIVES

In order to consider in more detail the impact of the metal used, formulations were prepared which all employed a common energetic matrix consisting of RDX passivated with wax and ammonium perchlorate, and only the metallic component was varied. In the first approach, the metallic fuel consisted exclusively of aluminum, magnesium, or beryllium. In the second study approach, aluminum served as the principal fuel and was partially replaced by small amounts of either beryllium or cerium. Using the same formulation strategy ensured that differences in measured performance could be attributed primarily to the properties of the metallic additives rather

Corresponding author: T. M. Klapötke;
E-mail: tmk@cup.uni-muenchen.de
DOI: 10.5937/OTEH26051K

than changes in the energetic matrix.

TKX-50 has recently been studied as a potential replacement for HMX in thermobaric formulations and showed extraordinary performance, introducing higher detonation properties in the observed mixtures [11-14].

3. OXIDIZERS

Tris(2,2,2-trinitroethyl) orthoformate (TNEF) (Fig. 2) is one of the most interesting examples of a purely CHNO-based oxidizer developed in recent years [13-17]. It could be shown that TNEF combines several properties that are rarely achieved simultaneously in organic oxidizers, namely a favorable oxygen balance, comparatively high density and a halogen-free composition that avoids chlorine-containing combustion products. In this respect, TNEF addresses key limitations associated with conventional oxidizers such as ammonium perchlorate. The significance of TNEF is further supported by its demonstrated applicability in practical formulations. Incorporation into nitrocellulose-, HTPB-, PDMS- and GAP-based systems has shown that TNEF is compatible with technologically relevant binders and can deliver promising thermal and combustion-related performance. Recent stability studies also indicate that TNEF possesses a more robust thermo-chemical profile than might be expected from its highly nitrated molecular structure. At the same time, the current state of research on TNEF is still at an early development stage. Comprehensive ballistic data, large-scale production routes, long-term aging studies, and full motor-level validation have so far not been realized. TNEF should therefore not be considered a replacement-ready oxidizer at present. Overall, TNEF is best regarded as a highly promising pre-technical oxidizer candidate. Its importance lies not only in its individual performance metrics, but also in demonstrating that purely organic CHNO compounds are realistic targets for use in propulsion applications. Further experimental and computational studies will determine whether this potential can be translated into application. Figure 1 shows chemical structure of TNEF molecule.

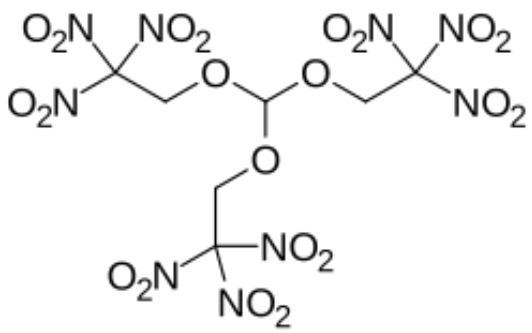


Figure 1. Structure of TNEF

The observed thermobaric composition with new oxidizer TNEF was made as cast-cured mixture with silicone binder, HMX/Al/TNEF/PDMS, and compared to a reference composition HMX/Al/AP/PDMS. They were labeled as TBX-1 (with AP) and TBX-2 (with TNEF).

4. CHARACTERIZATION OF THE THERMOBARIC FORMULATIONS

In order to characterize experimentally the explosive performance of selected thermobaric formulations, several techniques were used. For example, heat of detonation measurements were performed using bomb calorimetry under inert conditions to quantify the energy released during detonation. Quasi-static overpressure was measured inside a nitrogen-filled explosion chamber, allowing evaluation of energy release during the anaerobic combustion stage without interference from atmospheric oxygen. Finally, 500 g pressed charges were detonated in a semi-confined bunker where detonation velocity, peak blast pressure, and blast impulse were determined, thereby assessing the combined effects of anaerobic and aerobic combustion under realistic conditions. The experimental measurements were complemented by thermochemical calculations using the EXPLO5 code with the BKWs equation of state to compare measured behavior with equilibrium predictions.

5. RESULTS AND DISCUSSION

Calculated and measured heats of detonation for TBX formulations containing 60% RDX, 20% AP and 20% fuel (metal) are compared in Figure 2. LiF was included as an inert reference material. Here we may observe the effects of individual metal fuels incorporated in the reference thermobaric composition.

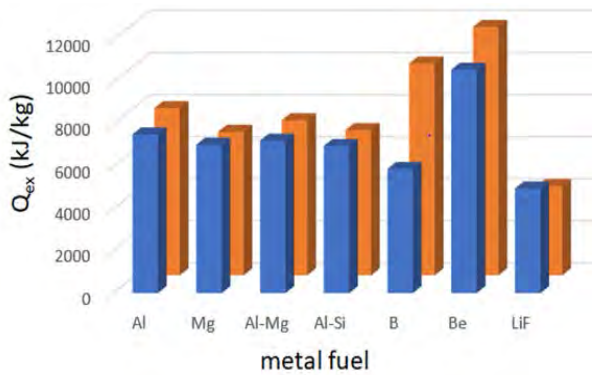


Figure 2. Calculated (orange) and measured (blue) heats of detonation for TBX formulations.

Direct comparison of the results obtained using aluminum, magnesium, and beryllium demonstrated that beryllium possesses a substantially greater ability to contribute energy during both stages of the thermobaric process. The beryllium-containing formulation produced the highest experimentally measured heat of detonation, exceeding the corresponding aluminum and magnesium formulations by more than forty and fifty percent, respectively.

Although equilibrium calculations predicted even greater energy release, the measured values indicate that reaction kinetics and incomplete oxidation limit full utilization of

beryllium under experimental conditions. Nevertheless, the energy contribution of beryllium remained significantly greater than that of conventional metallic fuels (Fig. 2).

Measurements in the nitrogen-filled chamber further demonstrated the exceptional activity of beryllium during anaerobic combustion. The composition containing beryllium generated the highest quasi-static overpressure and the largest TNT equivalent among the formulations investigated. These results show that beryllium reacts efficiently with the gaseous products formed immediately after detonation, even before atmospheric oxygen becomes available. Subsequent bunker experiments confirmed that this enhanced reactivity also translates into superior blast-wave characteristics under practical conditions, including increased peak overpressure and blast impulse. Overall, the data consistently indicate that beryllium is considerably more effective than either aluminum or magnesium in promoting energy release throughout the thermobaric event.

Building upon these findings, the second investigation explored whether smaller quantities of highly reactive metals could improve the thermobaric performance of aluminum-based formulations. Partial replacement of aluminum by only 4 wt% beryllium or cerium produced measurable improvements in explosive performance despite the comparatively small concentration added. Both modified formulations generated approximately ten percent higher quasi-static overpressures than the aluminum reference composition, while corresponding TNT equivalents derived from chamber measurements and blast-wave impulses also increased noticeably. These observations indicate that both additives (Be or Ce) accelerate or promote aluminum participation during the early stages of combustion, thereby improving overall energy utilization.

The behavior of beryllium and cerium, however, differs in important respects. Beryllium possesses an exceptionally high heat of combustion and therefore contributes both chemical activation and substantial additional energy release. Cerium, in contrast, has a considerably lower calorific value but exhibits high chemical reactivity toward oxidizing species generated during detonation. Consequently, cerium appears to function primarily as an ignition promoter that facilitates earlier aluminum oxidation rather than serving as a major energy source itself. Despite these different mechanisms, both additives successfully enhance the effectiveness of aluminum-containing thermobaric formulations.

Thermochemical calculations further support the experimental observations while also highlighting the importance of reaction kinetics. Equilibrium calculations generally predict greater energy release than is measured experimentally, indicating that complete oxidation of metallic particles is not achieved during the rapid expansion of detonation products. This discrepancy is particularly significant for beryllium because of its high oxygen demand and the limited time available for complete combustion. Nevertheless, the experimental results demonstrate that even partial reaction of beryllium provides substantial performance benefits owing to its high intrinsic energy content and rapid participation in post-detonation chemistry.

Regarding the thermobaric output of the formulations with new oxidizer, TNEF, the observed outputs are given graphically in Figures 3 and 4. It may be noted that TNEF provides equal to somewhat higher thermobaric output compared to AP. Being an eco-friendly, chlorine-free oxidizer, TNEF is a promising replacement for AP and other chlorinated oxidizers.

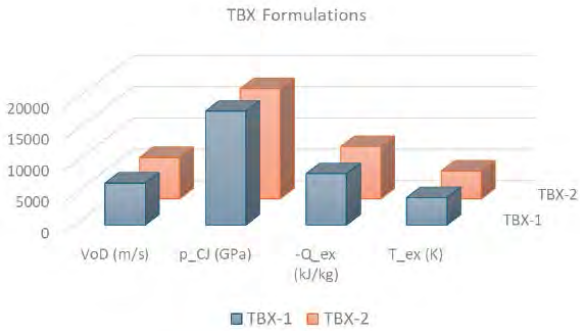


Figure 3. Graphical representation of the performance parameters of the TBX formulations TBX-1 (containing AP) and TBX-2 (containing TNEF instead of AP).

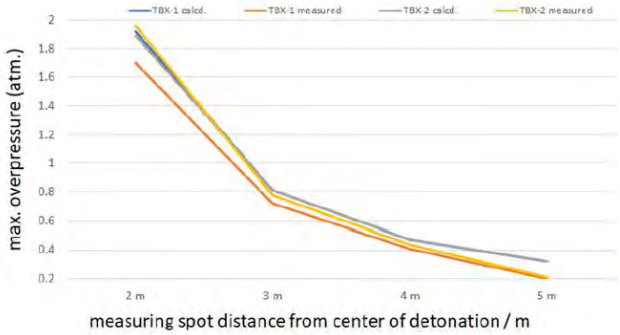


Figure 4. The maximum overpressures at different distances for TBX-1 (with AP) and TBX-2 (with TNEF) as a function of the distance

Replacing RDX in model thermobaric formulations by the highly promising secondary explosive TKX-50 was of particular interest since TKX-50 is a nitrogen-rich secondary explosive with the empirical formula $C_2H_8N_{10}O_4$ and therefore raised the question as to whether using a nitrogen-rich secondary explosive would reduce the quantity of oxidizer required in the TBX formulation, with the metal fuel potentially forming a metal nitride e.g. AlN (aluminum nitride) in the post-detonation products [11-12]. However, it was found that the heat of detonation of a 90/10 TKX-50/Al formulation was 90 J/g lower than the value for 100% TKX-50. It was found that in the post-detonation stage, Al reacts with the detonation products of TKX-50 contributing in this process 375 J/g. In an air-filled detonation chamber adding Al to TKX-50 results in a higher observed maximum temperature of the explosion products by 700 K, while this difference is reduced to 370 K in an argon-filled chamber. It was also found that the addition of Al under air-filled chamber conditions resulted in a much longer duration of time during which a temperature of at least 1600 K is observed. Interestingly

however, X-ray powder diffraction identified solid residue from the detonations performed under an Ar atmosphere as being predominantly $Al_{2.811}N_{0.435}O_{3.565}$ and not pure aluminum nitride, and under air-filled chamber conditions, essentially no aluminum oxynitride phase was observed. It should also be mentioned that TKX-50/ph/Al (70/3/27) shows a maximum overpressure value which is 30% higher than that of TKX-50ph (TKX-50/paraffin 97/3) in air atmosphere chamber conditions, which suggests that Al reacts with both oxygen from the air as well as the detonation products of nitrogen-rich TKX-50.

Based on these experimental observations, it was concluded that in this case, despite the higher concentration of nitrogen in the post-detonation products of nitrogen-rich TKX-50 in comparison with less nitrogen-rich secondary explosives, the burning of Al in nitrogen does not provide vast quantities of energy and aluminum oxides appear to be preferentially formed, even when the concentration of nitrogen in post-detonation products is considerably higher than that of oxygen.

6. CONCLUSION

Two promising approaches for improving thermobaric explosives have been described. Complete replacement of conventional metallic fuels by beryllium or, potentially, a beryllium-aluminum alloy, produces the greatest overall enhancement in detonation energy, quasi-static overpressure, and blast-wave performance. Alternatively, small additions of either beryllium or cerium to conventional aluminum formulations provide significant improvements while requiring only limited modification of existing compositions. The results show that careful selection and optimization of metallic fuel chemistry can substantially improve the efficiency of energy release during both the anaerobic and aerobic stages of thermobaric explosions. These findings provide valuable guidance for the future design of advanced thermobaric formulations in which both the energetic properties and the reaction kinetics of metallic additives are considered simultaneously. TKX-50 as a nitrogen-rich secondary explosive also significantly enhances thermobaric output of aluminized formulation in controlled conditions.

New oxidizer TNEF has shown potential as a chlorine-free replacement for AP on TBX formulations, providing even higher thermobaric output compared to reference AP-based TBX.

References

[1] Türker,L.: Thermobaric and enhanced blast explosives (TBX and EBX), Defence Technology, 12(6) (2016) 423-45.

[2] Trzcinski,W.,A., Maiz,L.: Thermobaric and Enhanced Blast Explosives - Properties and Testing Methods, Propellants, Explosives and Pyrotechnics, 35 (2010) 1-14.

[3] Simić,D.: Cast thermobaric PBX explosives - in Serbian, Scientific Technical Information - Monography Series vol. 53(2) ISSN 1820-3418,2016.

[4] Cudziło,S., Trzcinski,W.,A., Paszula,J., Szala,M., Chyłek,Z.: Performance of magnesium, Mg-Al alloy and silicon in thermobaric explosives a comparison to aluminium, Propellants, Explosives and Pyrotechnics, 45 (2020) 1-8.

[5] Yen,N.H., Wang,L.Y.: Reactive Metals in Explosives, Propellants Explosives and Pyrotechnics, 37 (2012) 143–55.

[6] Klapötke,T.,M., Cudziło,S., Trzciński,W.,A., Paszula,J.: Energy and Blast Performance of Beryllium in a Model Thermobaric Composition in Comparison with Aluminum and Magnesium, Defence Technology, 36 (2024) 13–19, https://doi.org/10.1016/j.dt.2024.02.011.

[7] Klapötke,T.,M., Cudziło,S., Trzciński,W.,A., Paszula,J.: Beryllium and Cerium Powders as Potential Activators of Aluminum in TBX Compositions, Propellants, Explosives and Pyrotechnics, (2025), https://doi.org/10.1002/prep.202400338.

[8] Klapötke,T.,M., Heidrich,J., Cudziło,S., Sućeska,M., Bajić,D., Bohn,M., Lechner,J.,T., Frehn,A., Reetz,B.: TBX Formulations with Alternative Metal Fuels, High Explosives and Oxidizers, 48th International Pyrotechnics Society Seminar (IPS 2025) September 15-19, 2025, Munich.

[9] Naglav,D., Buchner,M., Bendt,G., Kraus,F., Schulz,S.: Off the beaten track - a Hitchhiker's guide to beryllium chemistry, Angewandte Chemie International Edition, 55(36) (2016) 2-17, https://doi.org/10.1002/anie.201601809.

[10] Cudziło,S., Trzciński,W.A., Paszula,J., Klapötke,T.M.: Performance of Ammonia Borane Compared to Aluminum in Thermobaric Explosives, Propellants, Explosives and Pyrotechnics, (2026), in press: https://doi.org/10.1002/prep.70108.

[11] Klapötke,T.,M., Cudziło,S., Trzciński, W.,A.: An Answer to the Question about the Energetic Performance of TKX-50, Propellants Explosives Pyrotechnics, 47 (2022) e202100358, https://doi.org/10.1002/prep.202100358.

[12] Klapötke,T.,M., Cudziło,S., Trzciński,W.,A., Paszula,J., Bauer,L., Riedelsheimer,C. Lechner,J.: Performance of TKX-50 in Thermobaric Explosives, Propellants, Explosives and Pyrotechnics, 48 (2023) e202300010, DOI: 10.1002/prep.202300010.

[13] Klapötke,T.,M.: Chemistry of High-Energy Materials, 7th edn, deGruyter, Berlin/Boston, 2025.

[14] Klapötke,T.,M.: New Secondary Explosives and Oxidizers: TKX-50 and TNEF, Engineering Power, 20(1) (2025) 11–23: https://hrcak.srce.hr/file/493596

[15] Dosch,D.,E., Andrade,K., Klapötke,T.,M., Krumm,B.: An Optimized & Scaled-up Synthetic Procedure for Trinitroethyl Formate TNEF, Propellants, Explosives and Pyrotechnics, 46 (2021) 895-98.

[16] Schmidt,L., Thoma,M.,R.,H., Bohn,M., Klapötke,T.,M.: Tris(2,2,2-trinitroethyl) orthoformate (TNEF): A Promising Halogen-Free CHNO Oxidizer for Advanced Solid Propellant Applications, FirePhysChem, (2026) in press.

[17] Trache,D., Klapötke,T.,M., Maiz,L., Abd-Elghany,M., DeLuca,L.,T.: Recent Advances in New Oxidizers for Solid Rocket Propulsion, Green Chemistry, (2017) 4711-36.

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION V

Integrated sensor systems and robotic systems - ISSRS

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SIMULATION OF COMPLIANT ROBOTIC SYSTEM WITH A PARALLEL MECHANISM AND SCISSOR LIFT

NEMANJA TANASKOVIĆ
Institute BioSense, 21000 Novi Sad, nemanja.tanaskovic@biosense.rs

DAMIR KRKLJEŠ
Institute BioSense, 21000 Novi Sad, dkrkljes@biosense.rs

MIHAILO LAZAREVIĆ
Faculty of Mechanical Engineering, 11120 Belgrade, mlazarevic@mas.bg.ac.rs

***Abstract**: This paper presents a physics-based simulation of a modular compliant robotic system that couples a cable-driven parallel mechanism with a scissor lift extension stage. The mechanism employs three DC servo motors (rated at 490.3325 Ncm with a 25 mm pulley radius) and three antagonistic spring elements (stiffness k_s = 7878 N/m), arranged on a hexagonal base of radius R = 35 mm to achieve controlled tilt and vertical displacement of the end-effector platform. The simulation runs at 1 kHz within PyBullet and exposes a real-time graphical interface supporting both interactive and automated demonstration modes. Bézier curve interpolation generates smooth actuator trajectories between discrete waypoints. The simulation is coupled to a physical prototype via an STM32 Nucleo F401RE microcontroller through a UART serial link. Motor commands are encoded as PWM pulse widths in the range 500-2500 µs. The four-stage scissor lift provides vertical extension without disturbing the kinematic configuration of the parallel module. Simulation results demonstrate stable platform tracking across the full trajectory range under gravity loading, with no constraint violations observed during open-loop operation.*

***Keywords:** modular robotics, parallel mechanism, scissor lift, physics simulation, kinematic modelling, compliant actuation.*

1. INTRODUCTION

Parallel mechanisms offer a well-established combination of structural stiffness, high load-to-weight ratio, and predictable kinematic behaviour that serial-chain architectures cannot match at comparable actuator masses [1]–[3]. The most prominent is a Stewart-Gough platform, incorporating six independently extensible legs connecting a fixed base to a mobile platform, providing six degrees of freedom (DOF) pose control [4].

Many field applications, such as inspection, agriculture, wearable assist devices, and similar, require mechanisms that can accommodate unstructured contact, absorb disturbances, and be physically compact for transport, yet extendable to a useful working range on demand [5], [6]. Actuators compliance addresses the first two requirements. Whether implemented through series elasticity [7], antagonistic spring pairs [8], [9], or passive cable compliance, elastic elements are used to absorb impact energy, regularize force transmission, and decouple the controller from high-frequency mechanical resonances. Cables transmit only tension, removing the need for rigid link pairs, and the resulting low inertia enables fast reconfigurations of the workspace, giving cable-driven parallel robots (CDPRs) additional advantages over their rigid-link counterparts [5], [10]. The main constraint is that the cables must remain taut. This makes workspace definition and trajectory planning non-trivial tasks. However, for mechanisms operating about a nominal configuration, linear approximations of the inverse kinematics are tractable and have been validated experimentally [3], [11]. Scissor-type lifting mechanisms provide orthogonal capability and compact storage with almost linear vertical extension, described by a closed-form height-to-angle relationship [12], [13]. Integrating a scissor stage above a rotating platform module decouples the elevation positioning from the fine orientation, which simplifies both the kinematic model and the control hierarchy.

Modular architectures exploit these properties by stacking identical modules, each contributing additional DOFs without requiring a complete redesign of the controller [14], [15]. The platform described in this paper aims to achieve that type of design: a base module with three drive motors and three antagonistic springs provides tilt and limited heave, while a four-stage scissor lift provides the required vertical reach. Physics-based simulation at 1 kHz in PyBullet [16] closes the development loop before hardware deployment, with a—command-forwarding interface to an STM32 Nucleo [17] enabling open-loop system validation.

2. SYSTEM DESCRIPTION

The Mimi platform consists of two stackable modules and each module has two segments. Base segment that is doing

Corresponding author: N. Tanasković;
E-mail: nemanja.tanaskovic@biosense.rs
DOI: 10.5937/OTEH26052T

rotating motion and scissor mechanism segment that is doing translational motion along local z-axis. One such module is shown on Figure 1. Each base segment has three DC servo motors and three antagonistic spring elements that are mounted on a hexagonal base of radius $R = 35$ mm, with the actuators interleaved at 60° angular intervals. The motors drive the platform through cables, while the springs provide passive restoring forces that maintain cable tension and contribute stiffness. On top of each rotating segment, a four-stage scissor lift provides pure vertical extension along the module's local z-axis without altering the relative orientation between the scissor base and the platform above it. Two linear guide rails stabilize the scissor mechanism by reducing its lateral tilt. The scissor is driven by a separate spool and cable arrangement, decoupled from the tilt actuation. Stacking two modules with their respective scissor stages provides a system with independent tilt control at each level, enabling spatial configurations that a single-module design cannot achieve.

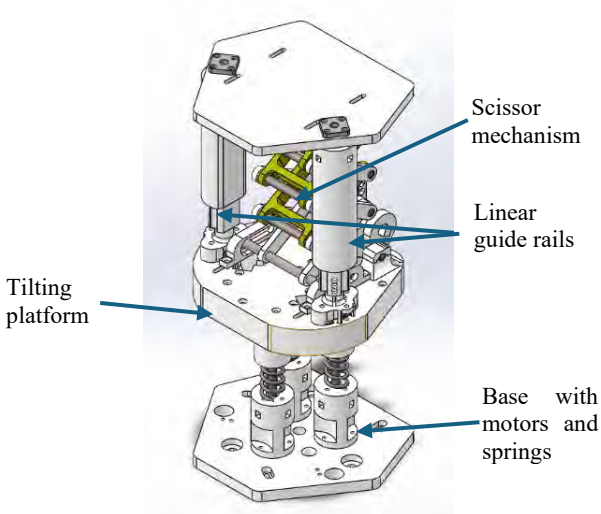


Figure 1 3D model of one full module of our robotic platform

3. KINEMATIC MODEL

3.1. Geometric Definition

The global fixed frame $F_o = \{O;\ \hat{e}_x, \hat{e}_y, \hat{e}_z\}$, whose origin coincides with the centre of the base plate shown in the Fig. 2. is used as a reference. Actuators are arranged in two interleaved sets at radius R. Main actuators, motors indexed $k \in \{1,2,3\}$ at angular positions:

$$\theta_{m,k} = \frac{2\pi(k-1)}{3}, \quad (1)$$

Error! No sequence specified.

with planar coordinates $p_{m,k} = R\{cos\theta_{m,k}, sin\theta_{m,k}, k\}^T$. Antagonistic springs are placed at $\theta_{c,k} = \theta_{m,k} + \frac{\pi}{3}$, with coordinates $p_{c,k}$ defined analogously.

The 60° offset ensures that each spring contact point lies at the bisector of two adjacent motor positions, thereby ensuring symetricall placement of motors and springs.

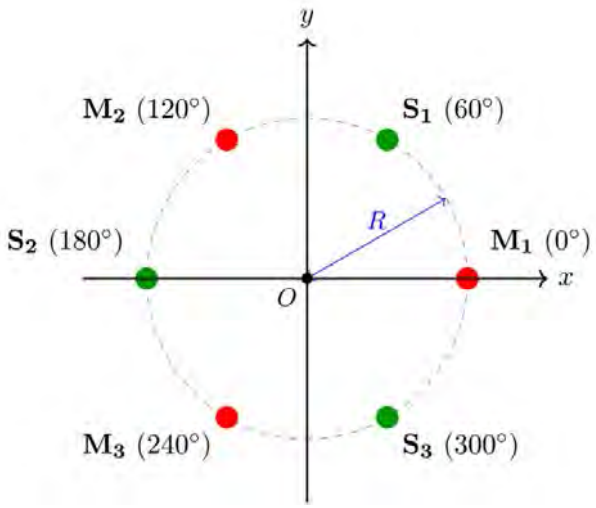


Figure 2. Motor (red) and spring (green) position on the base plate, with shown radius R

3.2. Actuation Kinematics

Let $\varphi(t) = \{\varphi_1, \varphi_2, \varphi_3\}^T$ be the motor input vector, with $\varphi_k > 0$ denoting cable shortening at motor k. The system operates at a nominal height z_0. The z-heights imposed at the main actuator contact points are:

$$Z_{m,k}(t) = z_0 - r_k \cdot \varphi_k(t), \quad (2)$$

while the antagonistic control actuators satisfy the constant-diameter constraint:

$$Z_{c,k}(t) = z_0 + r_k \cdot \varphi_{opp,k}(t), \quad (3)$$

where the opposing permutation is $\varphi_{opp,k} = \{\varphi_3, \varphi_1, \varphi_2\}^T$. Concretely, the spring at 60° opposes motor 3 at 240°; the spring at 180° opposes motor 1 at 0°; the spring at 300° opposes motor 2 at 120°. This pairing preserves platform planarity to a first order and is analogous to the constant-sum constraints exploited in antagonistic joint designs [8], [9].

3.3. Trajectory Generation: Constrained Bézier Interpolation

The system computes a travelling vector $V(\tau)$ that sweeps continuously around the hexagonal arrangement. The sweep parameter $\tau \in [0,1]$ is obtained by normalising the global azimuth angle within each 120° sector:

$$\tau = \frac{\theta_{global}\,(mod 120°)}{120°}. \quad (4)$$

Smooth interpolation between the discrete motor heights uses quadratic Bézier curves [18], [19]. The imposed constraint defines that the curve must pass exactly through the control actuator height $Z_{c,k}$ at the sector midpoint $\tau = 0.5$. For a standard quadratic Bézier with control points $P_{start}, P_{virt}, P_{end}$:

$$B(0.5) = 0.25P_{Start} + 0.5P_{virt} + 0.25P_{end} = P_{target}. \quad (5)$$

Solving for the virtual control point:

$$P_{virt} = 2Z_{c,k} - 0.5(P_{start} + P_{end}). \quad (6)$$

Substituting (6) back into (5), now evaluated at an arbitrary τ rather than at τ = 0.5, and collecting the terms in τ, the three scalar weight functions W_i are:

$$W_1(\tau) = 1 - 3\tau + 2\tau^2, \quad (7)$$

$$W_2(\tau) = 4\tau - 4\tau^2, \quad (8)$$

$$W_3(\tau) = 2\tau^2 - \tau, \quad (9)$$

The proposed weights satisfy $W_1(\tau) + W_2(\tau) + W_3(\tau) = 1$ for all τ. The partition of unity property, which ensures that the interpolated height is a convex combination of the actuator heights when $W_i \geq 0$ which is a property that bounds the travelling vector within the convex hull of the actuator values.

3.4. Per-Sector Height Equations

Substituting (2), (3), and (7)–(9) into the Bézier formula for each sector gives a new set of equations:

Sector 1 starts at motor 1, midpoint at spring 1, whose opposing motor is 3, and ends at motor 2:

$$V_{z,1}(\tau) = W_1(z_0 - r_k\varphi_1) + W_2(z_0 - r_k\varphi_3) + W_3(z_0 - r_k\varphi_2). \quad (10)$$

Sector 2 starts at motor 2, midpoint at spring 2, whose opposing motor is 1, and ends at motor 3:

$$V_{z,2}(\tau) = W_1(z_0 - r_k\varphi_2) + W_2(z_0 - r_k\varphi_1) + W_3(z_0 - r_k\varphi_3). \quad (11)$$

Sector 3 starts at motor 3, midpoint at spring 3, whose opposing motor is 2, and ends at motor 1:

$$V_{z,3}(\tau) = W_1(z_0 - r_k\varphi_3) + W_2(z_0 - r_k\varphi_2) + W_3(z_0 - r_k\varphi_1). \quad (12)$$

In all three sectors, z_0 factors out and the deviations from nominal collapse to a common matrix form. Let the sector index $s \in \{1, 2, 3\}$, then:

$$V_{z,s}(\tau) = z_0 + r_k\left[W_2(\tau)\varphi_{opp,s} - W_1(\tau)\varphi_s - W_3(\tau)\varphi_{s+1}\right], \quad (13)$$

which is the closed-form expression for the travelling vector height in any sector, parameterised entirely by the three motor inputs, φ, and the Bézier vectors, illustrated in Fig 3. When all motor inputs are equal, $V_{z,s}(\tau) = z_0$ for all τ, confirming that symmetric actuation produces pure heave without tilt.

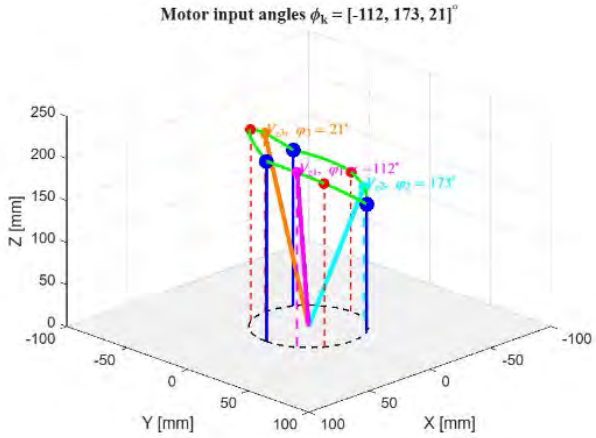


Figure 3 Bézier travelling vector (green) for φ = [−112, 173, 21]°. Motor contact heights (blue) and antagonistic spring heights (red) define the interpolation endpoints and midpoints in each of the three sectors

3.5. Scissor Lift Kinematics

The scissor lift, shown in Fig 4, consists of n = 4 identical stages, each built from crossed links of length $l = 0.08$ m. A spool of radius $r_{sc} = 0.0125$ m winds a control cable, and $l_0 = 0.05$ m is initial horizontal span of the scissor links at rest. The total lift height as a function of spool angle φ_{sc} is [12], [13]:

$$H(\varphi_{sc}) = n\sqrt{\left(\frac{l}{2}\right)^2 - \left(\frac{l_0 - r_{sc}\varphi_{sc}}{2}\right)^2} \quad (14)$$

The transmission Jacobian, sensitivity of lift height to spool angle, is:

$$\frac{dH}{d\varphi_{sc}} = -\frac{nr_{sc}(l_0 - r_{sc}\varphi_{sc})}{2\sqrt{l^2 - (l_0 - r_{sc}\varphi_{sc}^2)}} \quad (15)$$

The denominator in (15) tends to zero when the links approach full horizontal extension, which corresponds to a kinematic singularity of the scissor [20]. The operational range is chosen to remain well below this limit. Because the scissor operates along the platform's local z-axis, it is kinematically decoupled from the tilt mechanism.

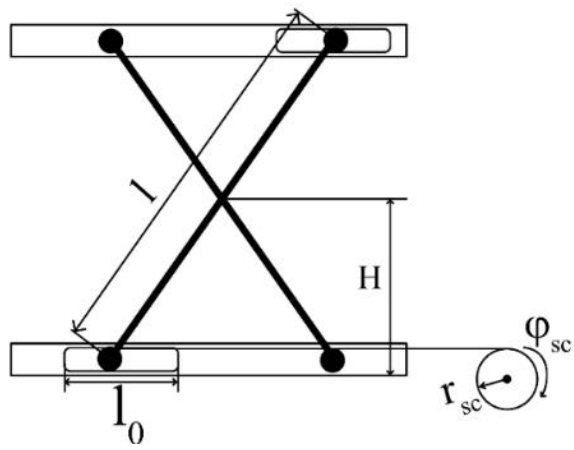


Figure 4. Schematic representation of one stage scissor mechanism

4. SIMULATION ENVIRONMENT

4.1. PyBullet Physics Engine

The simulation is implemented in Python using PyBullet [16], a physics engine that models rigid-body, contact, constraints, and gravitational loading. The comparison study of [21] confirms that PyBullet provides adequate fidelity for constrained rigid-body mechanisms of this class, with stable simulation at time steps of 1 ms. Gravity is set to $g = -9.81$ m/s²; the simulation time step is $\Delta t = 1$ ms, giving a 1 kHz update rate.

4.2. Actuator and Constraint Modelling

Each of the three motors and three springs is represented as a two-body assembly, where a cylindrical base body is fixed to the module frame, and a piston body is connected to it by a prismatic joint. The prismatic joint constrains relative motion to the actuator's local axis. A spherical joint connects the top of the piston to the platform, allowing free rotation at the attachment point and preventing the transmission of torques. This base prismatic piston spherical platform chain captures the essential constraint hierarchy of a cable-driven mechanism while remaining computationally tractable.

Each prismatic joint is actuated using a PD-based position control law that drives the piston toward its target height. The motor actuators receive target heights derived from (2), and the spring actuators receive target heights from (3). Gain values are tuned empirically and are not reported here, as they are simulation-specific heuristics rather than physical parameters of the mechanism. Figure 6 shows this chain for both actuator types. The motor and spring branches have an identical structural topology and differ only in the source equation used to determine the target height. Both branches converge at the spherical joint onto the shared platform body.

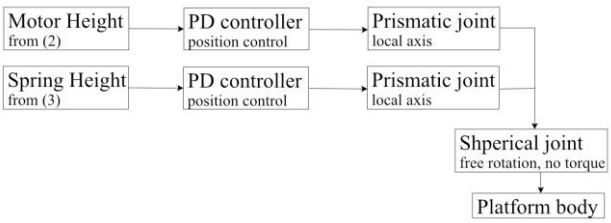


Figure 5. Block diagram of the actuator and constraint modelling chain for the motor and spring branches.

4.3. Module and Scissor Mechanism Modelling

Three custom classes were developed for Mimi platform, and are not part of the PyBullet library: Module, ScissorMechanism and SimulationController. Each class defines the structure and behaviour of its corresponding mechanical component. Object are instanitiated from these classes to construct the full simulation.

Object from the `Module` class encapsulates the platform body and six actuator objects, computing their attachment positions from the coordinate formulas. The ScissorMechanism object creates a prismatic constraint along the parent platform's local z-axis and computes the target piston position from (14) at each time step. The SimulationController is initiated at start up, stacks two modules, managing the kinematic chain from the ground to the top platform. Figure 7 show the resulting object hierarchy. The controller owns two Module instances, each of which contains its actuator objects, platform body, and a ScissorMechanism that recomputes its target piston position from (14) at every simulation step.

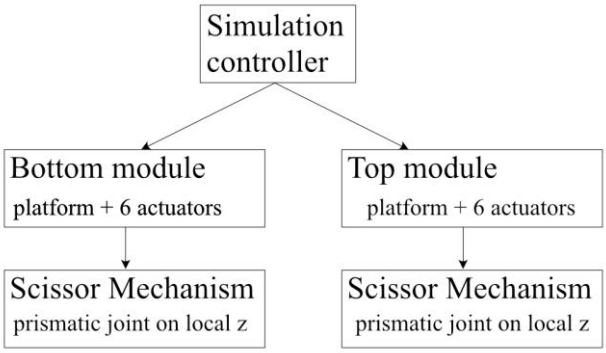


Figure 6. Object hierarchy of the Module and ScissorMechanism simulation classes.

4.4. Trajectory Visualisation and Demo Mode

A TrajectoryVisualizer object pre-computes the Bézier weight arrays (7)–(9) at initialisation and renders the travelling vector curve in real time by submitting line segments to the PyBullet debug drawing API. The curve updates every simulation step, so the user can observe the relationship between motor inputs and the interpolated trajectory continuously.

The demonstration mode drives the three motors with equal-frequency sinusoidal signals mutually offset by 120°, resulting in circular platform rotation. This mode is used for repeatable performance assessment without manual input, covering the full kinematic range of the platform within a defined time frame.

5. HARDWARE INTERFACE

5.1. Microcontroller and Communication Protocol

The DC servo motors on the physical prototype are controlled by an STM32 Nucleo-64 development board with STM32F401RE microcontroller[22]. The combination of a 32-bit ARM Cortex-M4 core, hardware PWM timers, and accessible UART peripherals makes it suitable for this purpose. During the simulation, the communication with the microcontroller was established over a UART serial link at 115200 baud with a DTR toggle reset sequence to synchronise firmware startup. Motor commands are transmitted as ASCII-encoded strings in the format (PWM_1,PWM_2,PWM_3) followed by a newline character, where each field is a pulse width in microseconds. This protocol is identical in structure to the approach used in [17] for robot open-loop validation, and its simplicity eliminates parsing overhead on the embedded side.

5.2. Motor Command Encoding

The servo motors interpret PWM pulse widths in the range *500–2500* µs as angles, with *1500* µs corresponding to the neutral ($\varphi = 0$) position. The mapping from motor angle to PWM pulse width is:

$$PWM[\mu s] = 1500 + \left(\frac{\varphi}{4.27} \cdot 500\right) \qquad (16)$$

The denominator 4.27, expressed in radians, corresponds to the full-scale angular excursion of the servo; the factor of 500 maps this range to the ±1000 µs deviation from centre. Equation (16) is computed in the `NucleoInterface` class and applied to all three motor outputs before transmission.

The STM32 firmware parses each incoming packet, extracts the three pulse width values, and writes them to the corresponding hardware timer channels. The 300 Hz output rate of the main simulation loop is well within the 50 Hz minimum update rate required by standard RC-type servo receivers, ensuring that the motors track the commanded trajectory without stale-data artefacts.

5.3. Simulation-Driven Hardware Validation

The hardware validation is achieved by sending the motor commands from the simulation environment to the Nucleo over UART, and the physical motors execute the commanded trajectories in open loop. The PyBullet simulation runs concurrently with the physical hardware, providing predicted states that can be compared against observed motion through external means. This architecture is sufficient for verifying that simulation-derived trajectories produce correct physical motion and that the communication protocol operates reliably. The internal encoders embedded in the servo motors are used exclusively by the motor driver for position holding and are not accessible to the simulation or the STM32 firmware. Consequently, no feedback path from hardware to simulation exists in the current implementation.

6. RESULTS AND DISCUSSION

6.1. Platform Tracking Under Gravity

The demonstration mode simulation was run for 60 seconds with gravity enabled and all six actuator constraints active. During the simulation, the platform completed multiple full revolutions of the travelling vector without any constraint violations or divergence of the prismatic joint targets. The weight functions (7)-(9) maintained their partition of unity property throughout, so the platform height remained within the envelope defined by the actuators' heights $Z_{m,k}$ and $Z_{c,k}$ at all times. Under asymmetric motor inputs (two motors at zero, one motor at positive deflection), the platform tilted by the expected angle, and the antagonistic opposing springs exhibited the predicted counter force, as evidenced by the piston position readings of the spring assembly.

To verify the simulation with respect to the analytical kinematic model, equation (13) was evaluated at a fixed interpolation parameter $\tau = 0.25$ across six, randomly chosen, motor input configurations spanning the operational range. At $\tau = 0.25$, the Bézier weight functions take fixed values $W_1 = 0.4375$, $W_2 = 0.7500$, $W_3 = -0.1250$, so the travelling vector height reduces to a weighted combination dominated by the antagonistic spring actuator. For each configuration, a settling time of 1 s was introduced before piston heights were sampled. The analytical value of $Vz_{,1}$ was then computed from equations (2), (3), and (13) evaluated at the set motor angles, assuming perfect motor tracking. The PyBullet value is the piston height measured after the settling period, reflecting the physical equilibrium reached by the PD controlled actuator under gravity loading. The kinematic model is exact by construction and any residual difference between the two columns in Table 1 originates entirely from the simulator. Results are summarized in Table 1. At the neutral configuration $\varphi = [0, 0, 0]$, the error was 0.03 mm, confirming that the simulation correctly encodes the kinematic model at rest. Across asymmetric configurations and the actuator heights in the range 160-245 mm, the mean absolute error was 2.6 mm, corresponding to approximately 1.2% relative error. This error is attributed to static deflections under gravity loading, as the feedforward gravity compensation was not included in the PD-controlled actuator model. PD controller gains were set empirically and are not guaranteed to minimize steady-state tracking error. The simulation introduces a bounded tracking offset that is consistent across all tested configurations.

Table 1. Analytical vs. PyBullet travelling vector height at $\tau = 0.25$, Sector 1 ($z_0 = 200$ mm, $r_k = 12.5$ mm).

φ [deg]	Analytical (mm)	PyBullet (mm)	Error (mm)
[−112, 173, 21]	217.33	214.32	3.01
[0, 0, 0]	200.00	199.97	0.03
[100, −100, 0]	189.09	187.84	1.25
[−100, 100, −100]	194.55	198.14	3.59
[50, 50, 50]	205.45	199.97	5.48
[200, −200, 100]	194.55	192.77	1.78

6.2. Scissor Lift Stability

The scissor mechanism was exercised over the range $\varphi_{sc} \in [0, 2.5]$ rad, producing height excursions consistent with (14). The Jacobian (15) indicates increasing force demand as the mechanism extends, which was reflected in the higher PD effort required near the top of the range. Across the full extension range, no lateral tilt of the scissor stage was observed, consistent with the constraining action of the linear guide rails described in Section 2.

6.3. Open Loop Hardware Validation

With the open-loop link active, the three servo motors on the physical prototype responded to commands generated by (16) within one communication cycle. The platform tilt on the physical prototype matched the simulation state within the angular resolution of visual inspection. The

primary source of discrepancy between simulation and hardware was a lag of approximately 100 ms at the start of each motion segment. Possible sources include communication latency, servo actuator dynamics, and current limiting under load. The PyBullet actuator model does not capture any of these effects. Quantifying the individual contributions requires dedicated instrumentation and is left for future work. Platform tilt angles were cross-checked against the simulation using a smartphone's built-in orientation sensor, and the readings were consistent with the simulated tilt throughout the trajectory.

6.4. Discussion

The simulation results confirm that the PyBullet rigid-body model captures the dominant kinematic behavior of the mechanism with adequate fidelity for open-loop validation. The mean absolute error of 2.6 mm across asymmetric configurations corresponds to approximately 1.2% of the operational range; the near-zero error at the neutral configuration confirms correct kinematic encoding, while the offset under asymmetric loading is consistent with the absence of feedforward gravity compensation in the PD actuator model. Feedforward gravity compensation is identified as the primary improvement to reduce this offset. Together, these results indicate that the simulation is a valid development tool for trajectory design and protocol verification. Future work will incorporate feedforward compensation based on (15) to reduce tracking error in this region.

7. CONCLUSION

This paper presented the physics-based simulation, kinematic model, and open-loop hardware validation of the Mimi modular, extensible and compliant robotic platform. The constrained quadratic Bézier interpolation scheme provides a safe travelling vector in a closed form that is directly implementable on embedded hardware. Simulation at 1 kHz in PyBullet confirmed stable tracking across the full kinematic range under gravity. Quantitative comparison of the travelling vector height against the analytical model across six input configurations at $\tau = 0.25$ yielded a mean absolute error of 2.6 mm (approximately 1.2% relative error). The remaining discrepancy was primarily attributed to static gravity sag in the PD-controlled actuator model in the absence of feedforward compensation. Open-loop hardware validation through an STM32 Nucleo-64 development board demonstrated reliable command forwarding at 115200 baud, with physical platform motion visually consistent with the simulation state. Future work will address multi-module tilt coupling at large deflections, and the integration of force sensing for contact-aware operation in agricultural and inspection tasks, as well as closed-loop HIL validation with explicit encoder or IMU feedback.

Acknowledgement

The authors acknowledge the support of colleagues at the BioSens Institute, University of Novi Sad. This research was supported by the research grant of the Serbian Ministry of Science, Technological Development and Innovations, grant No. 51-03-34/2026-03/200105 from 05.02.2026.

References

[1] H. D. Taghirad, Parallel Robots: Mechanics and Control. Boca Raton, FL: CRC Press, 2013.

[2] C. Gosselin and J. Angeles, “Singularity analysis of closed-loop kinematic chains,” IEEE Trans. Robot. Autom., vol. 6, no. 3, pp. 281–290, Jun. 1990.

[3] K. Liu, J. M. Fitzgerald, and F. L. Lewis, “Kinematic analysis of a Stewart platform manipulator,” IEEE Trans. Ind. Electron., vol. 40, no. 2, pp. 282–293, Apr. 1993.

[4] V. E. Gough and S. G. Whitehall, “Universal tyre test machine,” Proc. 9th Int. Conf. Automobile Div., Inst. Mech. Eng., pp. 392–394, 1956–1957.

[5] M. Zarebidoki, J. S. Dhupia, and W. Xu, “A review of cable-driven parallel robots,” IEEE Robot. Autom. Mag., vol. 29, no. 3, pp. 89–106, Sep. 2022.

[6] C. Gosselin et al., “AgroCableBot: Reconfigurable cable-driven parallel robot for greenhouse or urban farming automation,” Robotics, vol. 12, no. 6, Art. no. 165, Nov. 2023.

[7] G. A. Pratt and M. M. Williamson, “Series elastic actuators,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Pittsburgh, PA, USA, 1995, pp. 399–406.

[8] S. Wolf and G. Hirzinger, “A new variable stiffness design,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), Pasadena, CA, USA, 2008, pp. 1741–1746.

[9] Z. Liu, G. Xu, B. Zhao, and B. Li, “Design and analysis of a novel compliant actuator with variable stiffness,” IEEE Access, vol. 11, pp. 107350–107363, Sep. 2023.

[10] J.-P. Merlet, “Stability analysis of underconstrained cable-driven parallel robots,” IEEE Trans. Robot., vol. 29, no. 1, pp. 288–296, Feb. 2013.

[11] C. Gosselin, “Stiffness mapping for parallel manipulators,” IEEE Trans. Robot. Autom., vol. 6, no. 3, pp. 377–382, Jun. 1990.

[12] X. Li and K. Zhao, “Simulative calculation and optimal design of scissor lifting mechanism,” in Proc. IEEE Int. Conf. Mechatron. Autom. (ICMA), Changchun, China, 2009, pp. 2079–2083.

[13] V. Gurskyi, V. Kuzo, N. Maherus, and P. Krot, "Kinematic and force analysis of a scissor lift mechanism," Math. Models Eng., vol. 11, no. 4, pp. 230–242, Dec. 2025.

[14] M. Yim et al., “Modular self-reconfigurable robot systems,” IEEE Robot. Autom. Mag., vol. 14, no. 1, pp. 43–52, Mar. 2007.

[15] S. S. R. Chennareddy, A. Agrawal, and A. Karuppiah, “Modular self-reconfigurable robotic systems: A survey on hardware architectures,” J. Robot., vol. 2017, Art. no. 5013532, 2017.

[16] E. Coumans and Y. Bai, “PyBullet, a Python module for physics simulation,” http://pybullet.org, 2016–2021.

[17] A. Kumar, A. Mittal, R. Arya, A. Shah, S. Garg and R. Kumar, "Hardware in the loop based simulation of a robotic system with real time control and animation of working model," 2017 Int. Conf. Inventive Syst. Control (ICISC), Coimbatore, India, 2017, pp. 1-5.

[18] X. Zhao, Z. Cao, W. Geng, Y. Yu, M. Tan and X. Chen, "Path Planning of Manipulator Based on RRT-Connect and Bezier Curve," 2019 IEEE 9th Annu. Int. Conf. CYBER Technol. Autom. Control Intell. Syst. (CYBER), Suzhou, China, 2019, pp. 649-653.

[19] L. Zhang, L. Sun, S. Zhang and J. Liu, "Trajectory planning for an indoor mobile robot using quintic Bezier curves," 2015 IEEE Int. Conf. Robot. Biomimetics (ROBIO), Zhuhai, China, 2015, pp. 757-762.

[20] Teshigawara, Seiichi & Asada, Haruhiko. (2019). A Mobile Extendable Robot Arm: Singularity Analysis and Design.

[21] E. Berrocal, B. Sierra and H. Herrero, "Evaluating PyBullet and Isaac Sim in the Scope of Robotics and Reinforcement Learning," 2024 7th Iberian Robot. Conf. (ROBOT), Madrid, Spain, 2024, pp. 1-6.

[22] Z. Mingliang, G. Liquan, H. Chen, B. Benkun and L. Ziwei, "Design and implementation of control system for transport robot based on STM32 microcontroller," *2021 IEEE Int. Conf. Artif. Intell. Comput. Appl. (ICAICA)*, Dalian, China, 2021, pp. 466-469.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

REAL-TIME THERMAL IMAGE SYNTHESIS IN UNITY 3D VIA EMPIRICAL INFRARED ADAPTATION OF LIGHT SIMULATION MODELS

BOJAN ŠMIDLING
University of Belgrade, Institute Mihailo Pupin, Belgrade, bojan.smidling@pupin.rs

STEFAN STOJKOV
University of Belgrade, Institute Mihailo Pupin, Belgrade, stefan.stojkov@pupin.rs

Abstract: *This work introduces a method for synthesizing thermal imagery in real time within the Unity 3D game engine (Unity Engine/Unity) Editor. The approach builds on the engine's existing light simulation mechanisms, with adjustments that approximate behaviour in the infrared domain. While grounded in real life samples, the method is largely empirical, with calibration driven by the goal of achieving visually convincing results across complete scenes, including dynamic elements and motion. A central outcome is the ability to maintain real-time performance, on the order of 60 frames per second, while remaining sufficiently lightweight to integrate into pre-existing simulation systems developed in Unity. The method is evaluated through direct visual comparison with data captured from thermal cameras, alongside a user-based visual recognition task designed to probe perceived realism. Basic distinctions between material classes, such as metals, plastics, and fabrics, are represented, and the model accounts for ambient thermal contributions influenced by solar exposure and environmental temperature. The primary constraint lies in the preparatory effort required to define thermal characteristics for individual scene elements. Despite this, the approach proves to be a practical and adaptable solution for simulation contexts, particularly in defence and industrial domains, where real-time, visually credible thermal rendering is required without imposing significant computational burden.*

Keywords: *Synthetic infrared image generation, Real-time thermal rendering, Unity Engine, Infrared radiance modelling, Sensor-aware image synthesis.*

1. INTRODUCTION

Infrared (IR) and thermal imaging have moved well beyond their historical military niche into surveillance, pedestrian and vehicle detection, industrial inspection, search-and-rescue, and robotics operating under degraded visibility.

Despite its enormous potential, the development and deployment of IR systems face a significant bottleneck - the scarcity of suitable IR data. Training robust and reliable IR systems requires large, diverse, and high-quality datasets. However, obtaining such datasets in real world is challenging due to privacy concerns, logistical issues, and the difficulty in capturing a wide range of scenarios and conditions [1].

Synthetic IR image data generation is a natural response, following the precedent set by game-engine-based synthesis in the visible domain. This paper proposes a reproducible method for generating synthetic infrared images, and potentially image sequences, using Unity version 2022. The claim is a method proposal, not a fully validated thermal-sensing solution, and not a claim of complete radiometric accuracy. The primary contribution is synthetic infrared rendering in Unity. Although not primarily a dataset-generation tool, as used in [2], the method may later support such workflows. The paper is a starting point for a broader research direction, using simple, self-contained examples to demonstrate feasibility of the core rendering-repurposing strategy rather than delivering a large-scale benchmark.

Beyond dataset scarcity, a second motivation for engine-based infrared synthesis is controllability. Physical thermal data capture is constrained by weather, time of day, sensor availability, and scene access, whereas a synthetic pipeline allows systematic variation of object temperature, emissivity, geometry, and camera viewpoint within a single reproducible scene. This controllability is particularly relevant for generating edge cases, unusual object-background temperature contrasts, or repeated scenarios with only one parameter changed at a time, none of which are easily obtained from field data collection. The cost of measured data is high, and it is difficult to obtain infrared images in various scenarios [3]. The present paper does not exploit this controllability quantitatively, since it presents only two illustrative scenes, but the underlying method is structured so that later work could vary these parameters systematically once the core rendering approach has been established as viable.

A further consideration is computational accessibility. Physically based infrared rendering, if pursued from first principles, typically requires dedicated radiometric renderers, detailed material spectral databases, and

Corresponding author: B Šmidling
E-mail: bojan.smidling@pupin.rs

DOI: 10.5937/OTEH26053S

substantial computational resources to reach acceptable frame rates. By contrast, the approach in this paper deliberately trades physical completeness for accessibility, relying on an engine that is freely available, widely taught, and already familiar to many engineers working on simulation tooling. The priority is establishing the basic approach, rather than delivering the most physically complete implementation achievable.

A final point worth stating explicitly concerns reproducibility. Since the method relies solely on standard Unity components, mesh objects, the Standard Shader's emission channel, a terrain object, and a script computing thermal radiation using Stephan-Boltzmann law per object, it can in principle be reimplemented in any recent Unity version without specialized plugins or proprietary rendering extensions. This stands in contrast to some physically based infrared renderers, which may depend on custom radiometric pipelines that are harder for other researchers to reproduce or extend. Ease of reimplementation is treated here as a practical advantage of the proposed approach, separate from its physical fidelity, and is offered as an additional argument for adopting it as a starting point for further research rather than as a claim about its ultimate accuracy.

Section 2 reviews Unity's visible-range rendering pipeline. Section 3 develops the analogy between visible and infrared rendering. Section 4 details the proposed method. Section 5 presents illustrative examples, and Section 6 concludes with limitations and future work.

2. VISIBLE RANGE RENDERING OVERVIEW

Game engine renders a scene composed of mesh-based objects, each assigned a material referencing a shader that determines how the surface responds to lighting and is rasterized into pixel values. The built-in Standard Shader used in this work exposes albedo, metallic, smoothness, and an emission channel that allows a surface to appear self-illuminated independently of external scene lighting [4].

A virtual camera defines viewpoint, field of view, and output resolution [5]. For each frame, the engine determines visible geometry from the camera viewpoint, applies lighting calculations per the active rendering pipeline, and writes resulting colour values into a two-dimensional image buffer [6]. This mature, real-time-efficient stack is precisely what motivates repurposing it for infrared rendering rather than building a separate renderer from scratch.

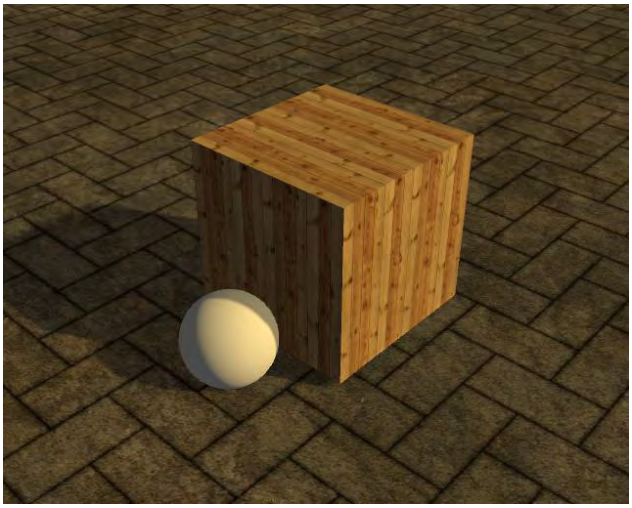


Figure 1. Example scene in visible range, concrete floor, wooden cube and brick sphere, illuminated by the Sun

3. CONNECTION AND ANALOGY OF VISIBLE AND IR RENDERING

The central idea is that the visible-range rendering stack need not be replaced, only repurposed, by changing what surface intensity represents. The same geometry, camera model, and largely the same material infrastructure are retained, but the value written to the output image is reinterpreted as an approximation of infrared emission rather than visible reflectance.

Unity does not natively render infrared radiation. Unity has no built-in concept of thermal emission or wavelength-dependent infrared material response. The method computes a simplified thermal value for each object outside Unity's native lighting model, then feeds that value into the emission property of the Standard Shader, reusing existing rasterization and display infrastructure. Emission functions purely as a rendering proxy, not as a physical model of infrared emission itself.

This distinction, between the physical thermal computation and the rendering mechanism used to display it, connects Section 2 to Section 4: once a thermal value is computed, it is injected into the existing visible-range pipeline through the emission channel, and the remainder of the pipeline proceeds exactly as for an ordinary scene. The method therefore inherits, essentially for free, Unity's existing support for real-time performance, dynamic elements, and animated characters, none of which need reimplementation.

4. METHOD

The proposed method is deliberately minimal in physical scope, covering surface temperature, surface emissivity, and a single spectral-band assumption framed as a long-wave infrared (LWIR) style approximation. Atmospheric transmission, reflected background radiation, detector calibration, sensor noise, optical blur, detector response, and angular or wavelength-dependent emissivity variation beyond the chosen band are intentionally omitted.

4.1. Thermal value computation

For a simplified broadband approximation, emitted radiant exitance is modelled using the Stefan-Boltzmann law:

$$M = \varepsilon\sigma T^4 \quad (1)$$

where M is emitted radiant exitance, ε is surface emissivity, σ is the Stefan-Boltzmann constant, and T is absolute surface temperature in Kelvin. Celsius inputs are converted to Kelvin before substitution.

Stefan-Boltzmann's law approximates total emitted thermal power integrated over all wavelengths, whereas real sensors measure band-limited radiance shaped by reflection, atmosphere, optics, and calibration. The present model is therefore a first-order thermal-appearance approximation, not a complete sensing model. Emissivity values are treated as approximate, band-specific constants per object in LWIR range, not universal material constants, since real emissivity depends on wavelength, angle, texture, and material state.

4.2. Object-level implementation

Each object is assumed to have a single, uniform temperature and emissivity across its visible surface, or each object is approximated to a diffuse reflector, which assumes the radiation reflection intensity of the target is equal in all directions, as stated in [7].. Shadows, ambient lighting, and post-processing are disabled or excluded where they would interfere with the simplified thermal logic, since they belong conceptually to the visible-light domain. In the implementation, every renderable object is parametrised by two scalar values: its surface temperature TTT (in °C, converted internally to K) and its surface emissivity ε, the latter taken from a long-wave infrared emissivity table [8]. In the example scene (**Figure 2**) the wooden cube is assigned $\varepsilon = 0.76$ at $T = 25$ °C (ambient), the concrete floor $\varepsilon = 0.90$ at $T = 25$ °C (ambient), and the brick sphere $\varepsilon = 0.94$ at $T = 40$ °C (self-heated), consistent with emissivity values reported for rough concrete, wood, and fine brick in LWIR .

The emitted radiant exitance M for each object is computed once from Equation 1, using its assigned temperature and emissivity, and serves as a purely numerical proxy for radiative power. This per-object thermal value is then mapped into a colour emission term through Equation 2:

$$EmissionColor = Color.white \cdot \frac{M(T,\varepsilon)}{M(Tmax,1)} \quad (2)$$

In the example configuration $Tmax = 100°C$, so all object emissions are normalised to the radiative output of an ideal blackbody at 100 °C. The resulting $EmissionColor$ is finally written into the _EmissionColor property of the Unity Standard Shader material for that object, ensuring that the visible emission intensity in the rendered frame is a monotonic function of the underlying thermal state while remaining independent of conventional lighting and shading passes.

Figure 2. Raw emission values image of wooden cube, concrete floor and brick sphere in LWIR mode of Unity

4.3. Terrain-level implementation

Outdoor scenes require a terrain surface, handled differently from discrete mesh objects since Unity's Diffuse Terrain shader does not directly expose an emissive parameter in the same way as the Unity Standard shader. The terrain shader takes a single RGB texture and tiles it repeatedly across the terrain. Our implementation takes a texture of grass, soil, and occasional stones, which is then converted into a thermal proxy by editing the artwork so that each material category is recoloured to a normalised grey level proportional to its emissivity, using the mapping:

$$Color_{part} = Color.white \cdot M_{part} \quad (3)$$

In effect, stone regions appear as the brightest tones, soils intermediate, and grass as the darkest, reflecting emissivity-like behaviour in the 0.90–0.98 range for typical soils and vegetation in LWIR, as stated in [8] and [9]. The resulting emissivity texture is then assigned as the albedo input to the Diffuse Terrain shader and treated as a spatially varying emissivity proxy over the terrain surface.

Since Diffuse Terrain provides no emission colour property, a dedicated directional light source is introduced to act as the terrain's thermal emission proxy. This light is restricted to the terrain via Unity's layer system, so that it illuminates only the terrain layer and no other scene objects, thereby isolating the terrain contribution from the rest of the lighting pipeline. For the terrain we set a uniform surface temperature $Tmax = 30°C$ and reuse Equation 2, to compute an emission colour $EmissionColor$. Instead of writing this value into Standard shader property _EmissionColor as in section 4.2., the $EmissionColor$ is applied directly to the C*olor* property of the terrain's dedicated directional light. The light is configured to mimic pure emission rather than conventional illumination by disabling shadows and reflections and aligning its direction to be perpendicular to the terrain surface, yielding a terrain contribution that is visually consistent with the rest of the thermal model while remaining conceptually separate from the visible-light domain.

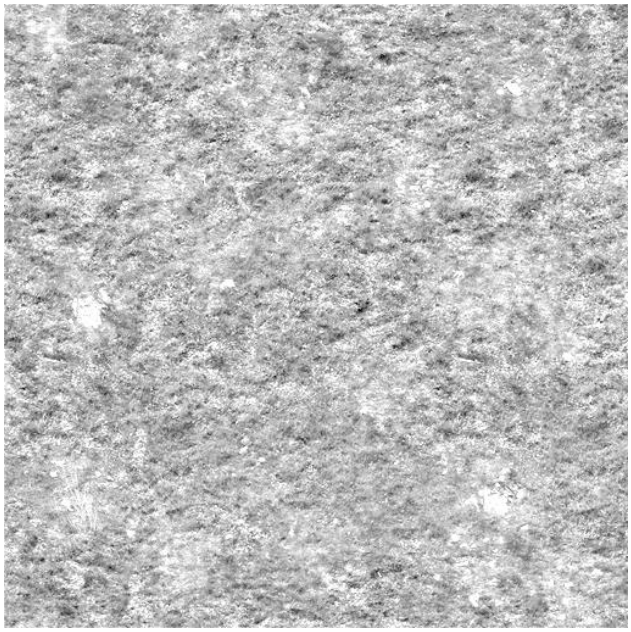


Figure 3. Emissivity texture of terrain used in example scene

4.4. Display mapping and thermal palettes

Once thermal values are computed for all objects and terrain, the scalar field is converted into a displayable image through a mapping stage. Pixel intensity is normalised using the minimum and maximum thermal values present in the current view frustum, so the coldest visible value maps to the darkest output and the hottest to the brightest. Because normalisation is per-frame, absolute brightness is not directly comparable across frames unless the underlying value range is also reported.

In the prototype implementation, the thermal field is rendered into a single render texture by Unity's camera, and a custom post-processing shader, *ReplaceColor*, is applied to this texture to perform the final colour mapping. For every pixel, *ReplaceColor* computes a normalised offset in the palette domain using its current blue-channel value (treated as the scalar thermal intensity) and the global minimum and maximum over the entire render texture,

$$offSetX = \frac{(inputTex2D.b * 255 - _minB)}{(_maxB - _minB)} \quad (4)$$

where *_minB* and *_maxB* are the minimum and maximum blue-channel values observed in the frame. *offSetX* acts as a lookup coordinate into a one-dimensional false-colour palette texture *_PalleteTex*. The final pixel colour is then obtained as

$$PixelColorValue = tex2Dlod(_PalleteTex, float4(offSetX, offSetY, 0, 0)) \quad (5)$$

where *offSetY* is a hard-coded offset selecting the active palette row (e.g. white-hot, red-hot, yellow-hot) within the palette texture. In this scheme the internal thermal output remains a single scalar intensity field, while the choice of palette texture and *offSetY* purely controls visualisation, following the behaviour of commercial thermal imaging systems where switching palettes changes only appearance and not temperature data.

The thermal output is represented internally as a single scalar intensity field. For display, it can be written as equal RGB values for grayscale output, or remapped through a false-colour palette such as white-hot, red-hot, or yellow-hot (Figure 4), following conventions common in commercial thermal software [10]. Palette choice affects only visualisation, not the underlying physical or computational content.

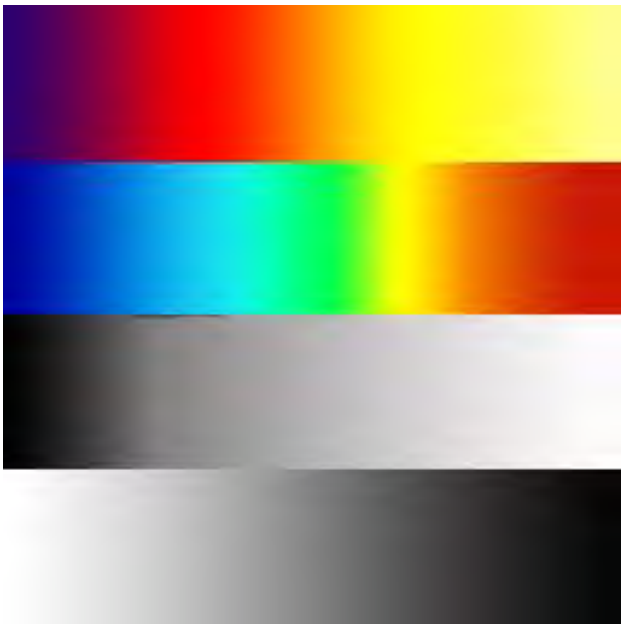


Figure 4. False colour thermal palettes, yellow hot - up, red hot - down

5. RESULTS

The examples below are illustrative demonstrations of feasibility, consistent with the paper's goal of establishing a starting point rather than a validated system. No quantitative sensor-level or task-based evaluation is performed here.

5.1. Uniform-material object scene

The first scene consists of simple geometric objects, each with a distinct material and a distinct temperature-emissivity pair, following Section 4.2. Concrete, wood, and brick surfaces were included, with object temperatures chosen to differ noticeably from one another. Under the normalisation procedure of Section 4.4., the rendered image shows visually distinguishable contrast between objects, driven entirely by Equation 1, without additional post-processing (Figure 5).

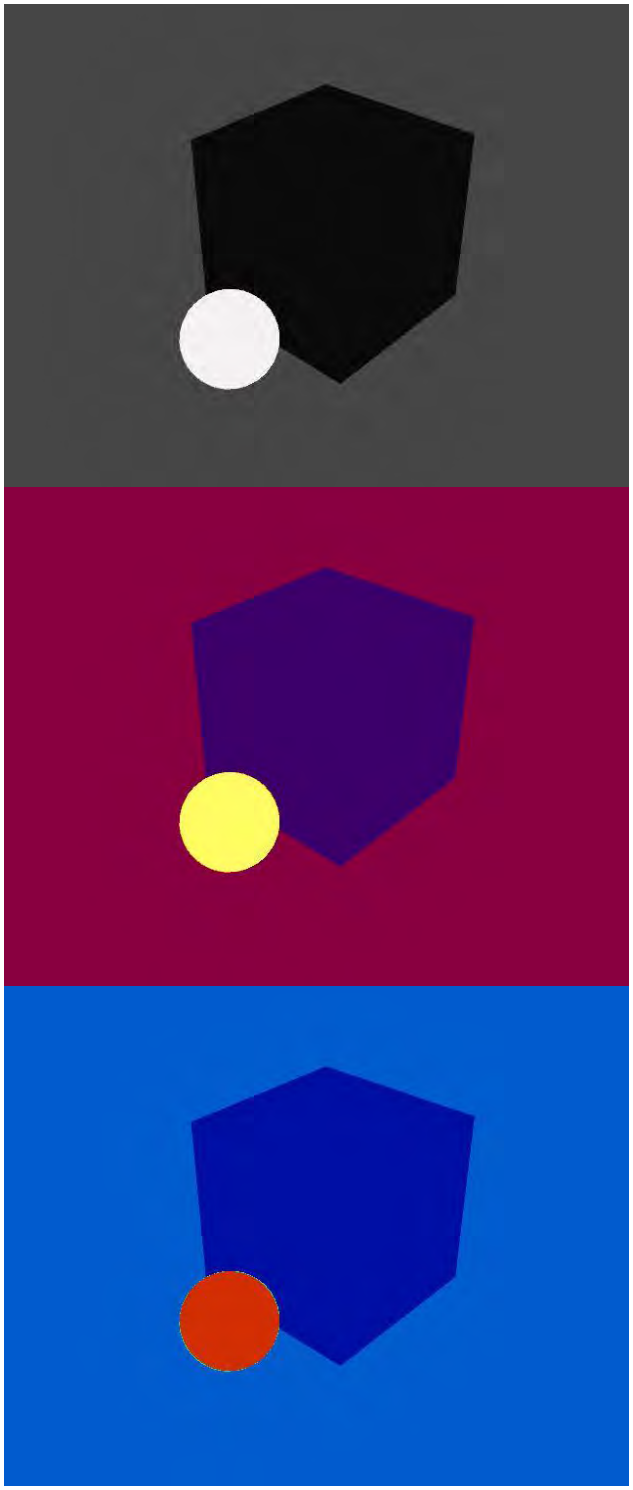


Figure 5. Synthetized LWIR image of example scene in White hot (top), Yellow hot (middle), and Red hot (bottom) colour modes, with normalization procedure

This scene isolates the core thermal computation and normalisation steps in a setting with no terrain or dynamic elements, so the basic mechanism's behaviour is not obscured by scene complexity.

5.2. Outdoor terrain scene with human figures

The second scene combines the object-level and terrain-level procedures of Sections 4.2 and 4.3. A real infrared photograph of soldiers in an open field served as a qualitative visual reference, not a ground-truth target for quantitative comparison. Soldier meshes were placed in a Unity terrain scene, assigned a common body and clothing temperature of approximately 36 degrees Celsius, converted to kelvin before use in Equation 1, together with approximate band-specific emissivity values for clothing and exposed skin of 0.88 (Fabric: Hessian, green) [8]. The terrain was treated with the thermal proxy texture and lighting workaround of Section 4.3., with grass, dirt, and stone assigned emissivity values in the 0.9 to 0.98 range and surface temperature of 30 degrees Celsius.

The rendered output shows human figures as distinctly brighter thermal forms against a cooler terrain background, qualitatively consistent with the reference photograph, where warm-bodied subjects stand out against cooler surroundings (Figure 6). This correspondence is offered strictly as evidence of plausibility, demonstrating that the method reproduces the basic figure-against-background contrast pattern typical of real thermal imagery of people outdoors, not that it reproduces exact radiometric values. Optional blur and synthetic sensor noise were identified as promising additions for improving perceptual resemblance to real thermal camera imagery, and are left for future work.

5.3. Interpretation and evaluation framing

Both scenes support plausibility only and are not sufficient to prove physical radiometric correctness. Appropriate evaluation depends on intended downstream use, and the present paper avoids committing to a single universal protocol that was not actually executed. For realism aimed at human observers, in simulation or training contexts, evaluation should rely on human judgement or structured realism review.

The image provided in [11] was used as a visual reference point (Figure 6). Top panel shows an LWIR image of several people on rough terrain, while the bottom panel shows a Unity-generated synthetic infrared image produced by the proposed method. Blur and noise were added in Photoshop to facilitate visual comparison with the reference image.

For algorithm testing or benchmarking, evaluation should rely on downstream task performance against an established benchmark protocol. Both strategies are left as future work.

6. CONCLUSION

This paper has proposed a deliberately simplified, first-order method for synthesizing infrared imagery in Unity 3D, and demonstrated through two simple examples that visible-range rendering infrastructure can be repurposed to produce structurally plausible thermal-like output. The method retains Unity's existing geometry, materials, and camera model unchanged, introducing a thermal computation stage based on a single-band (LWIR) Stefan-Boltzmann approximation, per-object temperature, and approximate emissivity, injected into the pipeline through the Standard Shader's emission property used strictly as a

rendering proxy.

The implementation is explicitly limited by object-wise uniform emissivity and temperature, a single spectral band (LWIR), and the omission of atmospheric transmission, reflected background radiation, detector calibration, sensor noise, optical blur, and angular or wavelength-dependent emissivity variation. Each of these omissions is identified as an addressable direction for future research.

Concrete extensions include material-class differentiation between metals, plastics, and fabrics through class-specific emissivity behaviour, incorporation of ambient thermal contributions from solar exposure and environmental temperature via Equation 1, and a move from qualitative plausibility toward quantitative evaluation through human-judgement studies or downstream task-based benchmarking. This work suggests that general-purpose real-time game engines are a viable and accessible foundation for synthetic infrared research, provided the physical scope and limitations of any such method are stated as clearly as they have been here.

Figure 6. LWIR image captured by IR camera (up) and LWIR generated image in Unity (down) with blur and noise added in post-process

Acknowledgment

Bojan Šmidling and Stefan Stojkov were supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, under Contract no. 451-03-33/2026-03/200034

References

[1] UPADHYAY, AVINASH & SHARMA, MANOJ & MUKHERJEE, PRERANA & SINGHAL, AMIT & LALL, BREJESH. (2025). A comprehensive survey on synthetic infrared image synthesis. Infrared Physics & Technology. 147. 105745. 10.1016/j.infrared.2025.105745.

[2] CARBONE, CARLOS & POTENA, CIRO & NARDI, DANIELE. (2020). Simulation of near Infrared Sensor in Unity for Plant-weed Segmentation Classification. 81-90. 10.5220/0009827900810090.

[3] ZHIJIAN H, BINGWEI H AND SHUJIN S (2022) An Infrared Sequence Image Generating Method for Target Detection and Tracking. Front. Comput. Neurosci. 16:930827. doi: 10.3389/fncom.2022.930827

[4] Unity Technologies, "Unity User Manual 2022.3 (LTS)", Accessed: Jul. 30, 2026. [Online]. Available: https://docs.unity3d.com/2022.3/Documentation/Manual/render-pipelines.html

[5] Unity Technologies, "Unity User Manual 2022.3 (LTS)", Accessed: Jul. 30, 2026. [Online]. Available: https://docs.unity3d.com/2022.3/Documentation/Manual/CamerasOverview.html

[6] ANDREW S. GLASSNER. 1994. Principles of Digital Image Synthesis. Morgan Kaufmann Publishers Inc.

[7] LI, BO & ZHAO, HUAI-CI & DU, MEI & LV, JINFENG. (2013). The Infrared Multispectral Image Simulation Research Based on 3D Scene. 10.2991/eeic-13.2013.32.

[8] ThermoWorks, "Emissivity Table", Accessed: Jul. 30, 2026. [Online]. Available: https://industrial-lab.thermoworks.com/blogs/learning-center/emissivity-table

[9] AN, NI & HEMMATI, SAHAR & CUI, YU-JUN. (2017). Assessment of the methods for determining net radiation on different time-scales of meteorological variables. Journal of Rock Mechanics and Geotechnical Engineering. 9. 10.1016/j.jrmge.2016.10.004.

[10] TOPDON Incorporated 2017, "How to Choose the Right Color Palette for Clearer Thermal Imaging", Accessed: Jul. 30, 2026. [Online]. Available: https://community.topdon.com/detail?id=892&type=1

[11] D. JEYABHARATHI, DR & DHARMA, DEJEY. (2018). Efficient background subtraction for thermal images using reflectional symmetry pattern (RSP). Multimedia Tools and Applications. 77. 10.1007/s11042-018-6220-1.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

IMPROVING DRONE DETECTION USING ACOUSTIC METHODS AND STATISTICAL FILTERING

MIODRAG VRAČAR
Military Technical Institute, Beograd, vracarmiodrag@mts.rs

STEVO VRAČAR
ITS Information Technology School, Beograd, stevo5817@its.edu.rs

RADIANA BRUSIN
Military Technical Institute, Beograd, radiana.brusin@mod.gov.rs

Abstract: *Acoustic methods for the detection of propeller-driven drones are used at relatively short distances, up to a few hundred meters. The drones are relatively weak sources of sound. The appearance of acoustic disturbances in the environment significantly reduces quality of the detection and localization drones by acoustic methods. Starting assumption, in resolving of this problem, is that acoustic ambient noise has a Gaussian natural distribution. A significant reduction in the influence of ambient noise was achieved by using the so-called statistical filter, which eliminates all contributions of ambient noise if their distribution is Gaussian. The function of filtering acoustic signal, that is, the filter itself, is realized as an independent module, that is, a processor that analyzes the acoustic signal using standard statistical methods, as well as higher order statistics. To check the quality of the filter itself it is used the five core quality factors: correlation coefficient, Normalized Root Mean Square Error (NRMSE), Signal-to-Error Ratio (SER), Maximum Absolute Error (MaxAE) and Estimated Delay (τ). The analyzed acoustic signals were generated in different flight mode of the drone, such as hovering, circular movement, left-right movement and others.*

Keywords: *drone, acoustic noise, statistical filter.*

1. INTRODUCTION

Acoustic detection of drones is crucial becausee it fills critical gaps left by radar and cameras against modern drone threats, especially in complex environments. There are more significant reasons for using acoustic methods such as its passive and steatlthy nature, identification of fiber optic drones, and in addition, these acoustic methods can act as a low-power trigger for turnig on radar and high resoluton cameras only after it hears a drone nearby.

There are different ways to do make signal useful and clear, such as filters and transforms. Nowadays exists and more advanced techniques like computer programs and smart networks. This paper explores a relatively little-used method for extracting useful signals from noise, which is statistical method based on higher order statistics. In adddition to the description of the filter itself, the paper presents its application in acoustics, seismics, hydroacoustics and infrasound analysis. Several tests were performed to obtain an evaluation of quality of the filter. The five core quality factors are used: correlation coefficient, Normalized Root Mean Square Error (NRMSE), Signal-to-Error Ratio (SER), Maximum Absolute Error (MaxAE) and Estimated Delay (τ). All of these methods help us to understand how useful is this filtering method for extracting a signal from noise.

Statistical filters use well-known statistical quantities, like mean and variance to process the signal. The mean filter, for example, replaces each data point with the average of neighboring points within a specific window. By averaging nearby values, high-frequency noise is smoothed out, and the signal is effectively de-noised, see [1]. Practical explanation of moving average filters with implementation examples and applications are given in the paper Smith, Steven W. [2]. Moving average is effective at removing high-frequency noise, but may blur sharp features or sudden changes in the signal. The choice of the window size is critical; a too small window might not remove enough noise, while a too large window can lead to loss of important signal details. There are many others filtering techniques. The Gaussian smoothing involves convolving the signal with a Gaussian kernel. The Gaussian function gives more weight to central points and less weight to distant points, see [3, 4].

Wavelet technique is used for de-noising signals by thresholding the wavelet coefficients. Wavelet transform decompose the signal and transform it into wavelet coefficients at different scales and positions. These coefficients represent the signal's details at various levels of time and frequency resolution [5, 6].

Classic comprehensive treatment of statistical filtering theory is well described in the book of the author Jazwinski, Andrew H. [7]. Modern Bayesian approaches including particle filters is described by Särkkä S. [8]. Comprehensive modern treatment of statistical filters including Kalman filters [9], particle filters, and their

Corresponding author: M. Vračar;
E-mail: vracarmiodrag@mts.rs
DOI: 10.5937/OTEH26054V

applications in tracking and estimation is described too. Particle filtering is based on sequential Monte Carlo methods for nonlinear/non-Gaussian problems such as target tracking, navigation, signal processing, finance, robotics, and others.

This paper presents the results of filtering acoustic signals, drone noise, using a statistical filter whose working principle is completely based on statistical methods. In the implementation of the filter, in addition to classical statistical procedures, such as the calculation of mean values, the so-called higher order statistics, i.e. the fourth order cumulants are used too. All the mentioned procedures are applied on signal segments – windows, with or without overlapping.

2 PRINCPLES OF THE FILTER DESIGN

Using the fourth-order cumulants (also called kurtosis in the statistical world) is a game-changer for machine noise analysis because they are "blind" to Gaussian noise and excel at detecting non-linear, impulsive signatures – exactly what failing or, standard mode of work, machine produce. Using them enables to:

- Suppress background noise (real-world acoustic sensors pick up wind, traffic, and electronic hiss-all Gaussian (random) noise. Fourth-order cumulants theoretically become zero for Gaussian noise, meaning they filter out the environment and leave only the machine unique "fingerprint".
- Reveals non-linear interactions. Machine parts, or drone propellers, interact in complex ways. The fourth-order cumulants capture phase coupling between frequencies, showing their interconnections, not just that they exist.
- Immune to synchronous noise. If propeller creates a loud, constant tone, the second-order methods see it as a big peak. But cumulants suppress these steady periodic noises, letting us to focus on the changing signatures.
- The method is robust to signal distortions. Mathematically it remains stable even if the microphone is cheap, or the signal clips slightly.

Relatively bad argument for using the fourth-order cumulants is that they require a lot of data (longer sampling time) to estimate accurately, and are computationally heavy. But for critical machinery – like drones, the early detection of its presence is worth the cost.

The flow chart diagram of the filter is shown in Figure 1. The process of statistical filtering is carried out in several steps. The first is segmentation with or without overlap. This is a significant considering that in order to successfully calculate the fourth-order cumulants, dynamically in time, a larger amount of data is required. The fourth-order cumulants are calculated over thus formed data segments, which contain information about the signatures of interest. Thus, this part of the analysis is responsible for extracting the frequencies of interest. In this way, the time-varying part of the signal was performed.

In practical applications DC signal component is also of interest. In order to preserve this component of the signal, a block named "Overall statistics of the signal segments" is created, see Fig. 1. Its basic role is to calculate the mean value of each signal segment. By adding these two branches, we get a filtered signal, or statistically filtered signal, which has information about both the DC component of the signal and all frequencies of interest of the signal.

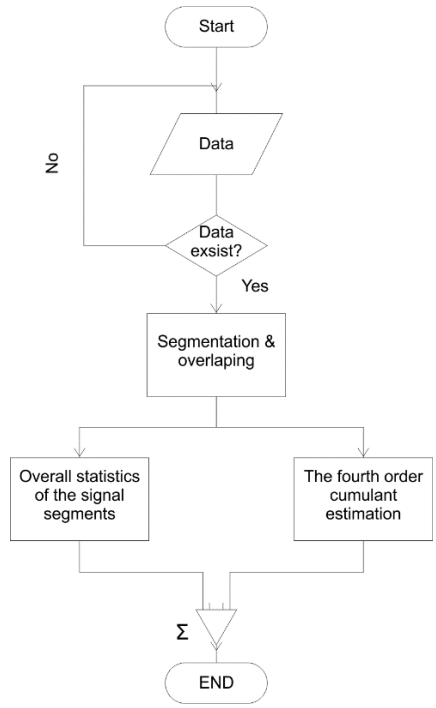


Figure 1. Flow chart diagram of the statistical filter algorithm.

The description of the statistical filter is given in the form of a block diagram, since its mathematical principles have been well known for a long time, and there is extensive literature on it. These signal analyses are very similar to the one who applies Short Time Fourier Transform (STFT), so it can be said that this approach opens up statistical domain in signal analysis. Statistical domain is just as equal the frequency domain considering that it closely connects them.

3. THE METRICS USED TO EVALUATE THE QUALITY OF THE FILTER

To check the quality of the filter itself it is used the five core quality factors: correlation coefficient, Normalized Root Mean Square Error (NRMSE), Signal-to-Error Ratio (SER), Maximum Absolute Error (MaxAE) and Estimated Delay (τ).

3.1 The Pearson correlation coefficient, ρ

The Pearson correlation coefficient, ρ, measures the linear relationship between source (X) and filtered (Y) signals:

$$\rho_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

where n is number of the samples.

3.2 Normalized Root Mean Square Error (NRMSE)

Normalized Root Mean Square Error (NRMSE) is defined by following equation:

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2}}{X_{\max} - X_{\min}}. \quad (2)$$

3.3 Signal-to-Error Ratio (SER)

Signal-to-Error Ratio (SER) measures how accurately the filter reproduces the desired frequency response.

$$SER = \frac{\text{Power of the output signal that passes correctly}}{\text{Power of the signal that is missshape}} \quad (3)$$

The "error" here isn't noise or distortion from external sources – it is the deviation of the real filter from the ideal mathematical model.

A high SER means the filter behaves almost exactly as deigned. A low SER means the filter introduces significant amplitude/phase errors.

If $H_{ideal}(f)$ = desired (ideal) frequency response, and $H_{real}(f)$ = actual measured frequency response, then the error in the frequency domain is:

$$E(f) = H_{real}(f) - H_{ideal}(f), \quad (3a)$$

and SER is:

$$SER = \frac{\int |H_{ideal}(f)|^2 df}{\int |H_{real}(f) - H_{ideal}(f)|^2 df} \quad (3b)$$

In decibels (most common):

$$SER_{dB} = 10 \cdot log_{10}\left(\frac{\int |H_{ideal}|^2}{\int |E|^2}\right) \quad (3c)$$

The main factors which contributes error in filters are: passband ripple, insufficient stopband attenuation, nonlinear phase, coefficient quantization (digital filters), and component tolerances (analog filters).

In case when we have input (X) and output (Y) amplitudes directly, the gain (G) of a system is defined as, $Y_{ideal\ output} = G \cdot X_{input}$, the error is the difference between the actual output and the ideal expected output. Power of the signal that is misshape is estimated starting from error amplitude, A_{error},, which is defined as follows:

$$A_{error} = Y_{actual\ output} - Y_{ideal\ output} \quad (3d)$$

If we don't know Y_{ideal_output}, SER_{dB}, is estimated, starting from Eq. (3), by following equation:

$$SER_{dB} = 10 \cdot log_{10} \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}. \quad (3e)$$

This definition of SER_{dB} is used in this paper.

3.4 Maximum Absolute Error (MaxAE)

Maximum Absolute Error (MaxAE) is estimated based on the following considerations:

$$max|X_{true} - X_{estimated}|$$

$$\text{or } \mathrm{MaxAE} = \max_{i=1,\ldots,n} |Y_i - X_i|. \quad (4)$$

MaxAE is the single largest deviation between the filter's actual response and the ideal response, measured across all frequencies of interest. The significance of this error is in fact that many standards specify a maximum allowed error, not an average.

3.5 Estimated Delay (τ)

The delay is estimated by finding the lag that maximizes cross-correlation:

$$\hat{\tau} = \arg\max_k |R_{XY}[k]| \quad (5)$$

where cross-correlation, R_{XY}, is estimated by

$$R_{XY}[k] = \sum_{i=1}^{n} X[i] \cdot Y[i+k], \quad (5a)$$

for $lag\ (k)$ (with appropriate zero-padding).

3.6 Signal power

The total signal power is the integral (or sum) of the squared magnitude of the frequency response:

$$P_{signal} = \int_{-\infty}^{\infty} |H(f)|^2 df$$

(6)

Power in decibels (dB):

$$P_{dB} = 10 \cdot log_{10}\left(\frac{P}{P_{ref}}\right) \quad (6a)$$

Common references: dBm (reference = 1 mW), dBW (reference = 1 W), and dB (no reference, just a ratio).

4. RESULTS OF THE ANALYSIS OF THE ACOUSTIC NOISE OF DRONES IN HOVERING MODE

The quality and capabilities of the acoustic filter were checked through the analysis of the acoustic noise of five types of drones that were in hovering mode. The results of the analysis are presented graphically, showing the real signal, the filtered signal, as well as the spectrogram of the original signal.

4.1.1 Drone DJI Matrice 210

DJI 210 Matrice 210 quadcopter is designed for industrial applications. Its total weight with batteries is 4.8 kg and overall dimensions 883 × 398 × 886 mm.

4.1.2 Drone DJI Mavic Pro

DJI Mavic Pro drone has total mass, when include gimbal cover, of 743 g, and folded dimensions 83 × 83 × 198 mm.
Maximal speed is 65 km/h.

4.1.3 Drone Parrot Anafi

Parrot Anafi is relatively small drone with overall

dimensions 244 × 67 × 65 mm and weight 320 g. Maximum horizontal speed is 15 m/s, and vertical speed 4 m/s.

4.1.4 Drone Yuneec Typhoon

Yuneec typhoon drone is hex copter (6 – rotors) designed for serial photography. Its weight is around 1695 g, and with maximum speed 13.5 m/s.

4.1.5 Drone DJI Phantom 4 Pro Plus

Total weight of this drone is 1388 g (battery and propellers are included). Maximal speed is 72 km/h, and diagonal size (excluding propellers) is 350 mm.

4.2 The quality of the filtrated data

It is possible to conclude subjectively about the quality of the performed filtering of the acoustic noise of drones by looking at the diagrams that show the source and filtered data, see Fig. 2. as their spectrograms, see Fig. 3, as well as objectively using parameters described in Section 3 of this paper. Both approaches have advantages and disadvanteges. The advantage of the first method is that it is fast and gives the first assessment of the acceptability of the solution. As for the second – analytical method, it is more objective because it is based on several statistical parameters.

Lookking at the graphs in Fig. 2, it can be concluded that the amplitudes of the filtered acoustic noisesignal are significantly smaller than the amplitudes of the raw data. This is consequence of the existence of a significant acousic noise that statistically has a Gaussian distribution.

Superior supression of Gaussian noise is performed, and therfore better SNR. The output have a significantly better SNR compared to linear filters (like FFT – based bandpass. The phase information is restored (better fidelity). Standard power – spectrum (second order) filters destroy phase information – they point out only what frequencies exist, but not when they occur. The fourth order cumulants preserve phase coupling. This means the output waveform preserves the shape of the impulses. The output is temporally accurate, which is critical for localizing drone by using sounds.

A 4th-order cumulant filter can isolate one source because it exploits statistical independence. The output is „cleaner“ in the sence that it belongs to only one physical source.

Linear filters struggle with „pink“ or „brown“ noise (common in industrial halls). The 4th-order filter outputs a signal free from this color, meaning the background sounds flat and uniform, making the fault signatures pop out visually and audibly.

If the acoustic noise itself is non-Gaussian (e.g., another drone or a hammering sound), the filter will not supress it. The output will be mix both sources.

In addition, because 4th-order cumulants amplify tails (large amplitudes), a single accidental loud bang (e.g., far and near explosions) can dominate the output temporarily ruining the signal quality. This unwanted propetry of the 4th-order cumulants can be overcome brcause the drone detection algorithm will omit these segments, bearing in mind that the drone represents a continuous source acustic noise.

In Table 1, the results of the statistical analysis of the acoustic signals of drones are systematized. Inspired by the multi-metric evaluation approaches found in literature [10], a novel aggregate Quality Factor (QF), see Eq. 6, defined as the weighted harmonic mean of individual normalized *RMSE*, *SER*, *MAE* ad estimated delay into a single percentage value, where the weights, w_i reflect the relative importance of each metric for the target application is proposed. In this case, the weighting factors were chosen as follows: w = [0.5, 0.2, 0.2, 0.1]. The greatest importance is given to the correlation factor, which is weighted by 0.5.

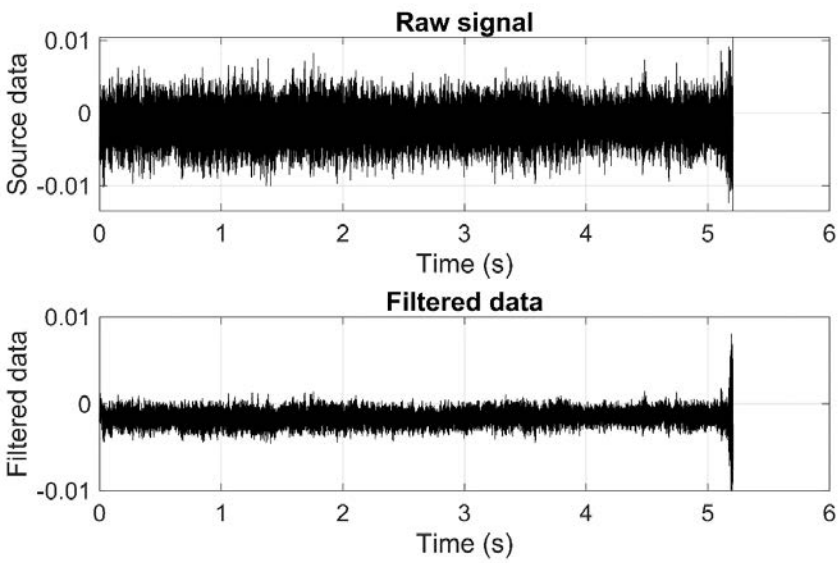


Figure 2. Raw and filtered data of the acoustic noise of DJI Matrice 210 drone in hovering mode.

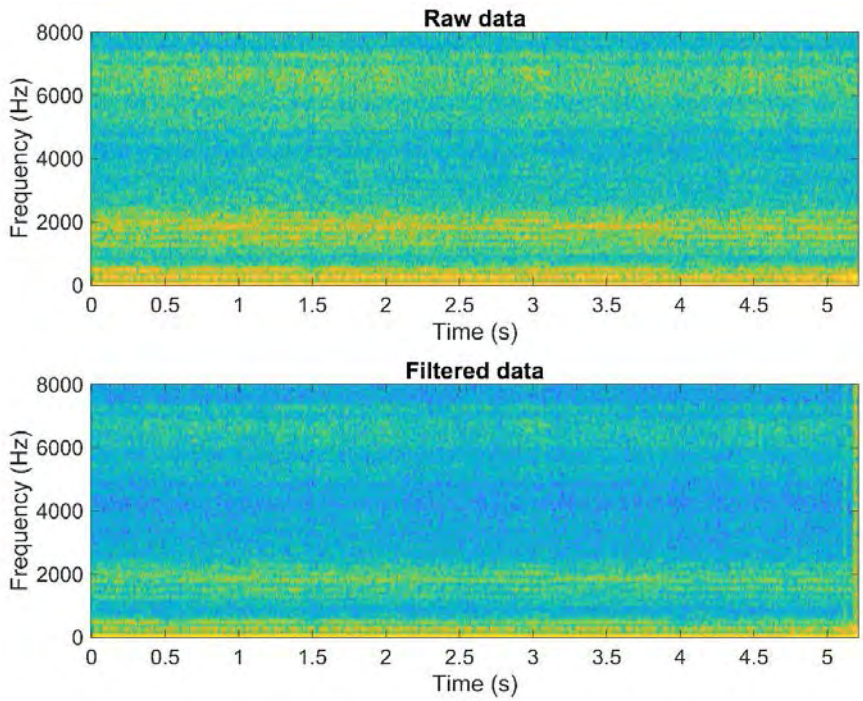


Figure 3. Spectrograms of raw and filtered data of acoustic noise of DJI Matrice 210 drone in hovering mode.

Table 1. Statistical parameters of the drone acoustic noise

+Statistical parameter that characterize source and filtered data	Type of drone				
	DJI Matrice 210	DJI Mavic Pro	Parrot Anafi	Phantom 4 Pro plus	Yuneec Typhoon
Correlation, ρ	0.9073	0.7248	0.8110	0.6960	0.8439
Error_mean	-0.0039	-0.0046	-0.0041	-0.0040	-0.0039
Error_std	0.0087	0.0068	0.0053	0.0069	0.0054
NRMSE	0.0378	0.0837	0.0807	0.0711	0.0574
SER_dB	13.4172	15.0861	16.0887	14.3544	15.7793
Max_Abs_Error (MaxAE)	0.1204	0.0619	0.0663	0.0595	0.0713
Delay_samples	27	33	30	45	46
Delay_seconds	0.0032	0.0040	0.0036	0.0054	0.0055
Signal_power	0.0020	0.0022	0.0018	0.0017	0.0017
Error_power	8.9689e-05	6.6790e-05	4.4406e-05	6.3199e-05	4.4115e-05
RMSE	0.0095	0.0082	0.0067	0.0079	0.0066
MAE	0.0070	0.0064	0.0052	0.0063	0.0052
Normalized_max_error	0.4798	0.6344	0.8034	0.5324	0.6158
Quality_score (%) (QF)	64.6052	55.2098	57.8653	57.0072	62.7528
Sampling_frequency (Hz)	8333	8333	8333	8333	8333

$$QF = w_1 \cdot \rho + w_2 \cdot (1 - \mathrm{NRMSE}) + w_3 \cdot \frac{\mathrm{SER}}{60} + w_4 \cdot \left(1 - \frac{\mathrm{MaxAE}}{0.1}\right) \tag{7}$$

By removing the Gaussian noise from the drone's acoustic noise, the result is that the spectrogram of the filtered signal, see Fig. 3, has become visually clearer, and the spectral components that originate from the operation of the drone's electric motors and its propellers are only clearly highlighted and dominate. The Gaussian noise (color disappeared in noisy areas) while preserving non-Gaussian signals (like machinery vibrations, or coms). The spectrogram's time-frequency resolution is unchanged, but the amplitude/color scale now represents non-Gaussian, not absolute power. Because 4th-order

filters are phase-blind, the spectrogram does not show phase coupling (bispectral peaks vanish).

As it well known the spectrogram is only interpretable if the signal is stationary over the window length; otherwise, time-varying cumulants distort the color scale. Therefore it is necessary to show the original spectrogram side-by-side, see Fig. 3, otherwise, the filtered version is meaningless without a baseline.

Finally, the 4th-order filter suppresses Gaussian noise but amplifies any non-Gaussian interference (e.g., clicks or transients), so it is necessary to inspect the spectrogram for artificial harmonic streaks caused by outliers, not real signals. These facts regarding filters based on 4th-order cumulants point to the need to integrate this filtering principle into algorithms that will identify these outliers in the signal end reduce their impact.

5. DISCUSION

It is well known that for acoustic based detection technology of the drones distinguishing is the distinctive features of acoustic patterns from ambient noise, and other acoustic sources. The noise spectrum consists of line spectra and their interconnecting components, where harmonic frequencies appear as distinct spectral lines. The primary goal is that acoustic signatures of drones become clearly differentiated from interference signals and ambient noise. Additional difficulty of the acoustic drone signature investigation is scenario un-homogeneity and ambient noise variability.

According obtained results, which are generated by application of statistical filtration method, which is based on different statistical procedures, such as classic statistical parameters and using higher order statistical methods on overlapped windows of the signal, relatively good results are performed. In all cases QF factor is greater from 50%, see Table 1. Signal to noise is significantly improved. Spectral lines that belongs to the drone activity are preserved as well. The raw acoustic data signal is converted into more appropriate form for further analytical purposes.

Implementation of the statistical filtration method in acoustic drone detection algorithms is significant preprocessing step, dedicated to extracting relevant features and reducing noise ambient interference.

Ambient noise compromise drone detection capabilities. Undoubtedly denoising plays an important role in drone acoustic detection. There are other approaches to reducing the impact of ambient noise such as methodology based on the linear shrinkage – subsequence projection – power spectral density filter (LSP) that employs covariance matrices and enhanced Wiener filtering, achieving effective noise suppression within a 135 m radius, see [11].

6. CONCLUSION

The paper shows the application of a statistical filter in the pre-processing of the raw acoustic signals, which originate from five types of drones in the hover mode. A significant increase in the signal-to-noise ratio was achieved, while retaining all the spectral components which originate from drone's activities.

The proposed statistical method of signal filtering largely eliminates the Gaussian noise, which makes up a significant part of the acoustic signal, since higher order statistics, i.e. cumulants of the fourth order, are implemented in the filter itself.

Nowadays deep learning techniques performed substantial advancements in drone acoustic detection. That deep learning architectures includes convolutional neural networks, recurrent neural networks, convolutional recurrent neural networks, and others. Used method for filtration of the raw acoustic noise drone's data can well serve to create better database for training advanced neural networks.

References

[1] OPPENHEIM, ALAN V. AND SCHAFER, RONALD W. AND BUCK, JOHN R.: *Discrete-Time Signal Processing*, Prentice Hall, Upper Saddle River, NJ, second edition, 176-182, (1999).

[2] SMITH, STEVEN W.: *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 277-284, (1997).

[3] YOUNG, RICHARD A.: *The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles*, General Motors Research Laboratories, Vol. GMR-4920, 1-38, (1987).

[4] GONZALEZ, RAFAEL C. AND WOODS, RICHARD E.: *Digital Image Processing*, Pearson, New York, 4th edition, 167-176, (2018).

[5] DONOHO, DAVID L. AND JOHNSTONE, IAIN M.: *De-noising by soft-thresholding*, IEEE Transactions on Information Theory, Vol. 41, No, 3, 613-627, (1995).

[6] DONOHO, DAVID L. AND JOHNSTONE, IAIN M.: *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, Vol. 81, No. 3, 425-455, (1994).

[7] JAZWINSKI, ANDREW H.: *Stochastic Processes and Filtering Theory*, Dover Publications, Dover reprint, New York, (2007).

[8] SÄRKKÄ, SIMO,: *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, (2013).

[9] JAZWINSKI, ANDREW H., *Stochastic Processes and Filtering Theory*, Dover Publications, Dover reprint, New York, (2007).

[10] ZIJAN ZHOU, YULI YANG, MENG MA AND BINGLI JIAC, *Distortion Reduction in Fractional Delay Filters*, IEEE Signal Processing Letters, PP (99):1-1, DOI:10.1109/lsp.2021.3064499, (2021).

[11] Y.SUN, Y. LIU, L. WANG, J.LI, J. WANG, A. ZHANG AND S. WANG, *Improved method for drone sound event detection system aiming at the impact of background noise and angle deviation*, Sens. Actuators, A 37, 115676, doi.org/10.1016/j.sna.2024.115676, (2024)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ADVANCED HYDROGEN DETECTION MATERIALS FOR CBRN APPLICATIONS: FROM METAL–ORGANIC FRAMEWORKS TO REAL-TIME THREAT ASSESSMENT

IVANA PEROVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 060 6690400, ivanaperovic@vin.bg.ac.rs

SNEŽANA BRKOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 063 8312423, snezana.miulovic@vin.bg.ac.rs

MINA SEOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 063 335658, mina050@vin.bg.ac.rs

STEFAN MITROVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 069 3639855, stefan.mitrovic@vin.bg.ac.rs

MILENA PIJOVIĆ RADOVANOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 069 5547598, pijovic@vin.bg.ac.rs

JELENA GEORGIJEVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 064 2757875, georgijevic@vin.bg.ac.rs

PETAR LAUŠEVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 069 5547598, petarl@vinca.rs

Abstract: *Hydrogen is not a conventional chemical, biological, radiological or nuclear warfare agent, yet its rapid diffusion, broad flammability range and very low ignition energy make an uncontrolled release a consequential secondary hazard at energy, transport, defence and research facilities. This critical review connects advanced hydrogen-sensitive materials with the operational requirements of CBRN reconnaissance and incident command. Palladium-based nanostructures remain the reference platform, while alloying, nanoscale confinement and porous supports mitigate hysteresis, poisoning and slow kinetics. Metal–organic and covalent organic frameworks serve as active phases, selective layers, nanoparticle hosts and precursors to porous oxides, but fast percent-level response and low-ppm detection must not be conflated. Recent machine-learning-assisted sensing improves compensation, early inference and source localization; evidence for autonomous discrimination of malicious releases from routine venting nevertheless remains limited. A defensible CBRN architecture therefore combines qualified sensors, environmental and process context, uncertainty-aware analytics, redundant communications and robotic standoff mapping while retaining deterministic alarm and shutdown logic.*

Keywords: *hydrogen detection; CBRN; metal–organic frameworks; threat assessment.*

Corresponding author: I. Perović;
E-mail: ivanaperovic@vin.bg.ac.rs

DOI: 10.5937/OTEH26055P

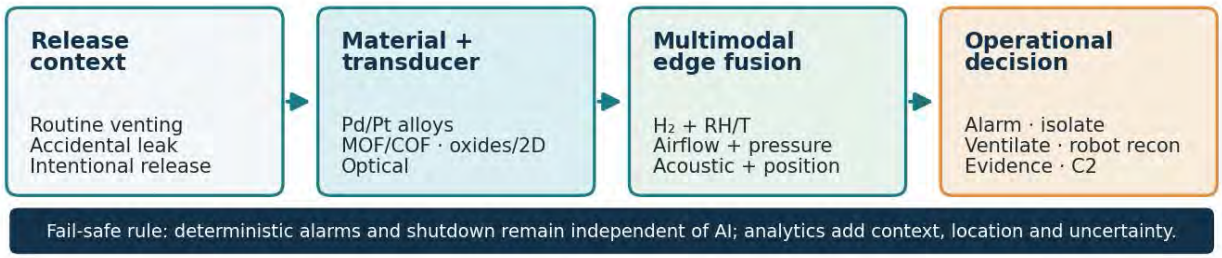


Figure 1. Material-to-decision architecture for hydrogen-aware CBRN monitoring (C2: command and control; RH: relative humidity)

Table 1. Selected advanced hydrogen-sensing results and their operational interpretation.

Platform	Mode / T	H_2 condition	Dynamic result	Robustness evidence	Operational caveat
Pd(II)@CrPy MOF [1]	Chemiresistive / RT	0.25–1 vol%	5–7 s response	Selective in tested gases	Fast, but not a ppm-level device
MOF-derived ZnO@GC [2]	Chemiresistive / 50 °C	Detectable to 10 ppm	Tens-of-seconds scale	Humidity/selectivity examined	Promising early warning; heating and recovery remain
Pd@amic-PT COF [3]	Chemiresistive / RT	1 vol%	5.3 / 3.5 s response/recovery	Reported stable vs RH and interferents	Percent-level benchmark; limited field ageing data
Pd@Pt/SnO_2 [4]	Chemiresistive / 25–50 °C	1000 ppm	1 s response; 9–31 s recovery	Laboratory selectivity tests	Excellent speed; poisoning/long-term duty require qualification
$Pd_{80}Co_{20}$ metasurface [5]	Optical / RT	LOD 2.5 ppm; t at ~1000 ppm	~0.85 s response	Ageing, temperature and interferents tested	Spark-free optical readout; nanofabrication/readout complexity
Ta–Pd thin film [6]	Optical / RT	Seven pressure decades	Sub-second with capping layer	Hysteresis suppressed by composition	Wide range; package and gas-matrix validation needed
Pd/WSe_2 [7]	Chemiresistive / 65 °C	LOD 10 ppb	10 min exposure protocol	40 cycles; 15–80% RH tested	Exceptional LOD does not imply sub-second alarm
Nanoplasmonic + NN [8]	Optical + AI / heated	LOD 100 ppm	Spectral inference	Synthetic air, 80% RH	Strong node-level evidence; not threat classification

LOD values, response definitions, flow rates and test atmospheres differ among studies; the table supports qualitative comparison, not direct ranking. RT: room temperature; NN: neural network.

1. INTRODUCTION

Hydrogen infrastructure is expanding from established chemical and refining sites to electrolyzers, storage modules, pipelines, refuelling stations, ports, vehicle depots and distributed backup-power installations. Hydrogen is colourless and odourless, diffuses rapidly through small openings and, in air, is flammable in range approximately 4–75 vol%. Its minimum ignition energy is about 0.017–0.02 mJ [9]. These properties demand detection well below the lower flammability limit (LFL). A commonly used safety alarm level is 25% of the LFL, corresponding to about 1 vol% H_2, although site-specific codes and hazard analyses govern the final set-point [10,11].

The CBRN relevance requires careful definition. Molecular hydrogen is not normally classified as a toxic chemical warfare agent. It becomes important to CBRN organizations because they routinely operate in uncertain, multi-hazard environments where an accidental or intentional hydrogen release may create fire, explosion, structural damage, access denial or secondary dispersion of hazardous material. The same response architecture like remote reconnaissance, hazard mapping, warning and reporting networks, protected decision-making and evidence preservation, can incorporate hydrogen channels alongside chemical, radiological, meteorological and imaging payloads. European capability priorities explicitly identify the integration of advanced CBRN sensors into small autonomous vehicles and drones [12].

This review evaluates advanced hydrogen-sensitive materials through that operational lens. The central question is not which material has the lowest laboratory LOD, but which combination of material, transducer, packaging, calibration, network and decision logic provides timely and trustworthy action under humidity, temperature swings, interferents, loss of communications and adversarial ambiguity.

2. REVIEW SCOPE AND PERFORMANCE REQUIREMENTS

A targeted critical review of peer-reviewed experimental papers, recent perspectives and official performance guidance available through July 2026 was undertaken. Priority was given to studies reporting concentration range, response/recovery behaviour, operating

temperature, humidity or interferent testing, stability, and integration with wireless or data-driven systems. Cross-study numerical ranking was avoided where authors used different definitions of response, LOD, carrier gas, flow, exposure time or baseline correction.

ISO 26142 defines performance and test requirements for stationary hydrogen detection apparatus, including accuracy, response time, stability, measuring range, selectivity and poisoning; it does not prescribe system installation or control logic [10]. For CBRN use, these product metrics are necessary but insufficient. An operational specification should also address: (i) t_{90} at safety-relevant concentrations rather than only at the LOD; (ii) recovery and latching behaviour after a large exposure; (iii) humidity, oxygen, CO, hydrocarbons, sulfur species, aerosols and decontamination chemicals; (iv) calibration transfer between nodes; (v) intrinsic or explosion safety; (vi) power autonomy and communications loss; and (vii) confidence reporting and raw-data retention. Long-term drift and poisoning often dominate real availability even when initial sensitivity is exceptional [13–16].

3. ADVANCED MATERIAL PLATFORMS

3.1 Palladium, platinum and alloy nanostructures

Palladium remains the reference hydrogen-recognition material. H_2 dissociates on Pd, atomic hydrogen enters interstitial sites and PdH_x formation changes resistance, work function, lattice dimensions and optical permittivity. Nanowires, nanoparticles, discontinuous films, nanogaps and plasmonic arrays amplify one or more of these changes [14,15]. The same Pd–H phase transition can generate hysteresis, mechanical fatigue and slow recovery. Surface adsorption of CO, sulfur-containing species, water or hydrocarbons blocks catalytic sites, while oxygen changes surface kinetics. Therefore, selectivity is not an intrinsic constant: it is a function of alloy composition, morphology, capping layer, operating temperature and gas matrix.

Alloying Pd with Au, Ag, Cu, Co, Ni or Pt can suppress the α/β-hydride transition, accelerate kinetics or reduce poisoning. Luong et al. engineered Pd nano-patchy metasurfaces and reported a 2.5 ppm LOD; $Pd_{80}Co_{20}$ responded in about 0.85 s at approximately 1 mbar H_2 with optical readout [5]. Ta–Pd thin films achieved a hysteresis-free optical response over seven pressure decades and sub-second dynamics when paired with a suitable capping layer [6]. Bimetallic Pd@Pt/SnO_2 combined dissociation catalysis with semiconductor resistance modulation, producing a 1 s response to 1000 ppm H_2 at 25–50 °C [4]. A 2025 catalytic-plasmonic Pt nanoparticle design further illustrates the move toward high-humidity optical operation [17]. Optical transduction is attractive in explosive zones because the sensing head can be electrically passive and interrogated remotely, but the optical source, spectrum acquisition, fouling and alignment still require engineering qualification.

Reliability research is shifting from one-time performance to maintainability. Kim et al. traced degradation of a wireless Pd-nanowire node to carbon-containing surface contamination and restored nearly 100% of initial performance using a 200 °C, 10 min thermal refresh after two months of contamination [18]. This is operationally important: self-diagnostics and scheduled regeneration may be more valuable than a marginal improvement in pristine LOD. However, regeneration consumes energy and may be incompatible with an active hazardous atmosphere, so it must be inhibited during an alarm and managed as a maintenance function.

3.2 MOFs, COFs and framework-derived materials

Porous frameworks offer molecular-scale control over pore size, surface chemistry and guest–host interactions. Their role in hydrogen sensing falls into four categories that should be reported separately: a conductive MOF/COF as the active chemiresistor; a porous selective overlayer on another transducer; a scaffold that disperses Pd/Pt and mediates charge transfer; or a sacrificial precursor that is converted into a porous oxide/carbon composite [19,20]. Labeling all four simply as “MOF sensors” obscures the actual recognition and transduction mechanisms.

A direct example is Pd(II)-decorated conductive chromium–pyrazine MOF (Pd(II)@CrPy), which responded selectively in 5–7 s, but over 0.25–1 vol% H_2—not at 1–10 ppm [1]. By contrast, MOF-derived ZnO@graphitic-carbon microcubes detected down to 10 ppm and were evaluated under humid conditions, but their dynamic times were on the order of tens of seconds and they required mild heating [2]. Montoro et al. showed that calcination temperature and precursor identity control the selectivity of MOF-derived CuO, NiO and ZnO, with ZnO favouring H_2 [21]. These studies support MOFs as design tools for porosity and dispersion, but not a universal claim of simultaneously achieving single-ppm detection and an 8 s response.

COFs provide greater covalent stability and π-conjugated charge-transport pathways. Thokala et al. used a redox-active polyimide COF to reduce and anchor Pd nanoparticles without an external reducing agent. The resulting Pd@amic-PT chemiresistor operated at room temperature and showed 5.3/3.5 s response/recovery at 1 vol% H_2, with reported tolerance to humidity and interferents [3]. This is a compelling architecture for low-power nodes, yet percent-level tests do not establish low-ppm performance, multi-year stability or resistance to CBRN decontaminants. Framework stability against water, acids/bases, radiation, mechanical shock and thermal cycling must be assessed after device fabrication, not inferred from powder characterization alone.

3.3 Metal oxides, two-dimensional materials and hybrid interfaces

Semiconducting oxides such as SnO_2, ZnO, WO_3, In_2O_3 and TiO_2 are mature, inexpensive and readily

miniaturized, but conventional devices often require heating and respond to many reducing gases [13,14]. Noble-metal sensitization lowers the operating temperature and accelerates charge transfer, while porous or heterojunction structures increase active area. Pd@Pt/SnO_2 is an instructive example of catalytic and electronic sensitization acting cooperatively [4]. For battery-powered robotics, the duty cycle of microheaters—not only peak heater power—should be included in the mission energy budget.

Two-dimensional semiconductors can translate local catalytic events into large conductance changes. A Pd-decorated WSe_2 device detected 10 ppb H_2 at 65 °C and retained a measurable response at 80% RH [7]. Yet the protocol used 10 min H_2 exposures and 5 min purges. The result demonstrates remarkable analytical sensitivity, but it is not evidence of a 10 ppb, sub-second safety alarm. This distinction between LOD and safety response time is essential when selecting a CBRN payload.

3.4 Transduction, packaging and mission-oriented selection

Recognition chemistry cannot be evaluated independently of transduction. Chemiresistive devices are compact and compatible with low-cost electronics, but the complete node must control contact resistance, baseline drift, heater duty cycle and electromagnetic susceptibility. Optical devices permit an electrically passive sensing head and can exploit wavelength, intensity or spectral-shape changes; their penalties are source and detector complexity, optical fouling and alignment. Catalytic and thermal-conductivity devices provide robust complementary channels at higher concentrations, whereas electrochemical cells may offer useful low-concentration response but introduce electrolyte life, orientation and replacement constraints [13,14,16]. A heterogeneous node that combines physically different channels is often more defensible than multiple copies of a single high-sensitivity material because common-mode interferents are less likely to produce concordant signals.

Packaging determines the exposure actually seen by the material. Hydrophobic membranes, particulate filters, flame arrestors and selective overlayers can protect a sensing film, yet every added diffusion barrier changes response and recovery. Sample pumps improve placement flexibility but add power, acoustic signature, tubing adsorption and a new failure mode. Consequently, system response should be measured from gas arrival at the device inlet to authenticated alarm output, including transport, filtering, analogue settling, digital filtering, edge inference and communication latency. Calibration records should include gas matrix, humidity, temperature, pressure, flow and certified-mixture uncertainty. For fielded CBRN equipment, built-in test must distinguish an intact electronic channel from a chemically responsive sensing surface; electrical continuity alone cannot reveal catalyst poisoning or blocked pores.

Mission profile should therefore govern material choice. Fixed protection at a hydrogen plant prioritizes rapid and repeatable behaviour near safety-relevant fractions of the LFL, low false-alarm burden and maintainability. Portable early-warning instruments may value ppm capability, low mass and low standby power, but still require an unambiguous high-level alarm channel. A moving robot needs response and recovery times commensurate with vehicle speed and sampling height; otherwise its concentration map is spatially displaced by sensor memory. Long-duration unattended nodes prioritize drift compensation, energy per measurement and remote health reporting. No single LOD, response time or sensitivity value captures these trade-offs, and technology rankings should be stated for a defined concept of operations rather than as universal material league tables.

4. FROM SENSOR SIGNAL TO REAL-TIME THREAT ASSESSMENT

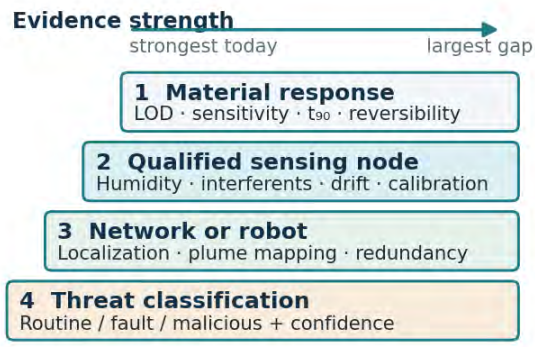


Figure 2. Evidence ladder: strong material results do not automatically validate an autonomous CBRN decision.

Machine learning is most defensible when it solves a clearly bounded sensor problem. Tomeček et al. processed the full spectrum of a heated nanoplasmonic sensor with dense neural networks or transformers and obtained a 100 ppm LOD in synthetic air at 80% RH [8]. Martvall et al. used a long short-term transformer ensemble model to predict the eventual plasmonic response in oxygen-free inert gas, accelerating the reported response by up to a factor of 40 while outputting prediction uncertainty [22]. These studies demonstrate humidity compensation, use of rich spectral features and early prediction; they do not demonstrate attribution of malicious intent.

At network level, data fusion can detect, quantify and localize releases. A recent hydrogen-refuelling-station study fused simulated concentration histories with wind speed and direction in a temporal convolutional network. After sensor-layout optimization, the model reported a mean source-location error of 0.54 m using 60 s sequences from 135 virtual sensors [23]. A four-node Zigbee system with a hydrogen/CO/alkane/smog array and an optimized back-propagation network reported correlation coefficients above 0.997 and mean relative errors below 0.25% in laboratory tests [24]. Both are promising, but the former depends on CFD-generated data and the latter used a small controlled dataset. Domain shift—from one geometry, sensor lot, weather regime or release type to another—remains a central safety risk.

A real-time CBRN assessment should therefore separate four outputs: detection ("is H_2 present?"), quantification ("what concentration and rate of change?"), localization ("where is the likely source and plume?"), and attribution ("is the pattern consistent with normal venting, a component fault or a deliberate act?"). Attribution requires contextual features that a gas sensor cannot provide: process state, valve commands, maintenance schedule, access-control events, video/acoustic evidence, wind field, node agreement and event history. The model should expose confidence and out-of-distribution status. A high hydrogen alarm must trigger deterministic ventilation, isolation or evacuation regardless of the classifier's interpretation; AI can prioritize investigation but must not suppress a safety trip.

Validation must match the level of the claimed decision. Randomly splitting adjacent time samples between training and test sets overstates generalization because neighbouring samples share the same release, weather and sensor history. A stronger design withholds complete incidents, facilities, sensor lots and environmental regimes, then reports sensitivity, specificity, false alarms per operating hour, time-to-decision and calibration of confidence. Simulation and CFD are valuable for generating rare geometries and hazardous release cases, but final performance should be anchored to blind physical trials. Threat attribution should be treated as human-supervised evidence fusion: the algorithm may rank hypotheses and identify inconsistent observations, while the incident commander retains authority and the system preserves raw data, model version and rationale for after-action review.

5. CBRN DEPLOYMENT ARCHITECTURE

For fixed critical infrastructure, a layered network is preferable to reliance on a single "best" material. Fast safety nodes should be located at credible release points and high-level accumulation zones, supported by environmental sensors and at least one complementary transduction principle. Optical or passive heads are advantageous in ignition-sensitive zones; chemiresistive or electrochemical nodes provide compact redundancy. Gateways should time-synchronize data, buffer raw signals during communication loss and use authenticated, encrypted links. Alarm thresholds, maintenance state and model version must be auditable.

For reconnaissance, a UGV can carry a calibrated H_2 sensor, RH/temperature, airflow, pressure/acoustic leak indicators, visible/thermal cameras and SLAM positioning. A UAV can rapidly survey elevated or inaccessible points, but rotor wash distorts light-gas plumes and must be included in sampling strategy. Mobile robots have already demonstrated gas-source localization by combining concentration sensing with LIDAR-based mapping in hazardous indoor simulations [25], and defence planning identifies autonomous platforms as a route to CBRN standoff sensing [12]. The hydrogen-specific next step is controlled, full-scale validation against known releases, obstacles, ventilation states and robot trajectories.

Operational reporting should preserve three layers: an immediate safety message (concentration, alarm state and protective action), a tactical map (estimated source, plume, uncertainty and sensor health), and a forensic record (time-stamped raw data, calibration, platform position and chain of custody). This structure supports both emergency response and later investigation without overloading the first responder with material-level sensor details.

6. EVIDENCE GAPS AND RESEARCH PRIORITIES

First, standardized tests must move beyond dry single-gas chambers. Candidate materials should be evaluated at 0.1–2 vol% H_2 for safety dynamics and at lower concentrations for early warning, across realistic RH/temperature cycles, oxygen levels, CO, methane, ammonia, sulfur compounds, aerosols and decontamination vapours. Report t_{10}, t_{50} and t_{90}, recovery after overload, false-alarm rate, drift and energy per measurement. Second, framework-based devices require packaged-device durability data: adhesion, pore blockage, radiation response, vibration and cleanability. Third, datasets for AI must include different sensor lots, facilities, seasons, routine vents, faults and controlled adversarial scenarios. External validation should be performed before claiming threat classification.

Fourth, sensor placement and mobile sampling should be co-designed with plume physics. CFD and digital twins are useful for rare-event coverage, but models must be anchored to full-scale releases and quantify uncertainty. Fifth, networks need graceful degradation: local alarms during gateway loss, self-test, cross-node plausibility checks, signed firmware and a manual fallback. Finally, field trials should use operational metrics—time to reliable alarm, source-location error, false dispatches, time saved in exclusion-zone mapping and responder dose/risk reduction—rather than laboratory sensitivity alone.

The literature supports fast MOF/COF-based devices and separate low-ppm MOF-derived devices, but not a general "1–10 ppm in 8 s" MOF benchmark. It supports AI-enhanced hydrogen quantification and localization, but no directly comparable peer-reviewed field study located in this review established 95% threat-versus-routine classification within 30 s or a 60% response-time improvement across transport hubs and critical infrastructure. These values are better presented as prospective validation targets until a traceable dataset and protocol are available.

7. CONCLUSIONS

Advanced hydrogen detection is progressing from sensitive materials toward context-aware safety systems. Pd-based nanostructures and alloys currently offer the strongest combination of rapid recognition, tunability and compatible electrical or optical transduction. MOFs and

COFs add selective transport, high-area supports and designable electronic pathways, while MOF-derived oxides bridge framework chemistry with established semiconductor sensing. The decisive CBRN advance, however, will not come from one material. It will come from qualified heterogeneous nodes, multimodal fusion, reliable communications, robotic standoff mapping and transparent decision logic. AI is already useful for compensation, accelerated inference and localization; malicious-intent classification remains an open, evidence-limited task. A safety-first architecture should keep threshold alarms and shutdown deterministic, using AI to add localization, context and uncertainty—not to replace fail-safe protection.

References

1. M. E. DMello et al., "Pd(II) decorated conductive two-dimensional chromium–pyrazine metal–organic framework for rapid detection of hydrogen," Int. J. Hydrogen Energy 47 (2022) 9477–9483. https://doi.org/10.1016/j.ijhydene.2022.01.003
2. A. Sharma et al., "Metal organic framework-derived ZnO@GC nanoarchitecture as an effective hydrogen gas sensor with improved selectivity and gas response," ACS Appl. Mater. Interfaces 14 (2022) 44516–44526. https://doi.org/10.1021/acsami.2c10706
3. N. Thokala, K. Vankayala, K. Basavaiah and S. B. Kalidindi, "Spontaneously decorated palladium nanoparticles on redox active covalent organic framework for chemiresistive hydrogen gas sensing," Int. J. Hydrogen Energy 81 (2024) 270–279. https://doi.org/10.1016/j.ijhydene.2024.07.230
4. X. Meng, M. Bi and W. Gao, "Rapid response hydrogen sensor based on Pd@Pt/SnO_2 hybrids at near-ambient temperature," Sens. Actuators B Chem. 370 (2022) 132406. https://doi.org/10.1016/j.snb.2022.132406
5. H. M. Luong et al., "Sub-second and ppm-level optical sensing of hydrogen using templated control of nano-hydride geometry and composition," Nat. Commun. 12 (2021) 2414. https://doi.org/10.1038/s41467-021-22697-w
6. L. J. Bannenberg, H. Schreuders and B. Dam, "Tantalum–palladium: Hysteresis-free optical hydrogen sensor over 7 orders of magnitude in pressure with sub-second response," Adv. Funct. Mater. 31 (2021) 2010483. https://doi.org/10.1002/adfm.202010483
7. X. He et al., "A resistivity-type palladium decorated WSe_2 device for ultralow concentration hydrogen detection," Adv. Compos. Hybrid Mater. (2025). https://doi.org/10.1007/s42114-025-01447-2
8. D. Tomeček et al., "Neural network enabled nanoplasmonic hydrogen sensors with 100 ppm limit of detection in humid air," Nat. Commun. 15 (2024) 1208. https://doi.org/10.1038/s41467-024-45484-9
9. U.S. Department of Energy, "Hydrogen Safety," Hydrogen, Fuel Cells & Infrastructure Technologies Program, 2004. https://www1.eere.energy.gov/hydrogenandfuelcells/pdfs/h2_safety_fsheet.pdf
10. ISO 26142:2010, Hydrogen detection apparatus—Stationary applications, International Organization for Standardization, 2010 (reaffirmed 2021).
11. W. J. Buttner, M. B. Post, R. Burgess and C. Rivkin, "An overview of hydrogen safety sensors and requirements," Int. J. Hydrogen Energy 36 (2011) 2462–2470. https://doi.org/10.1016/j.ijhydene.2010.04.176
12. European Defence Agency, "The 2023 EU Capability Development Priorities," 2023. https://eda.europa.eu/docs/default-source/brochures/qu-03-23-421-en-n-web.pdf
13. T. Hübert, L. Boon-Brett, G. Black and U. Banach, "Hydrogen sensors—A review," Sens. Actuators B Chem. 157 (2011) 329–352. https://doi.org/10.1016/j.snb.2011.04.070
14. W.-T. Koo et al., "Chemiresistive hydrogen sensors: Fundamentals, recent advances, and challenges," ACS Nano 14 (2020) 14284–14322. https://doi.org/10.1021/acsnano.0c05307
15. I. Darmadi, F. A. A. Nugroho and C. Langhammer, "High-performance nanostructured palladium-based hydrogen sensors—Current limitations and strategies for their mitigation," ACS Sens. 5 (2020) 3306–3327. https://doi.org/10.1021/acssensors.0c02019
16. T. M. Swager et al., "Critical sensing modalities for hydrogen: Technical needs and status of the field to support a changing energy landscape," ACS Sens. 9 (2024) 2205–2227. https://doi.org/10.1021/acssensors.4c00251
17. A. Theodoridis, C. Andersson, S. Nilsson, J. Fritzsche and C. Langhammer, "A catalytic-plasmonic Pt nanoparticle sensor for hydrogen detection in high-humidity environments," ACS Sens. 10 (2025) 8983–8994. https://doi.org/10.1021/acssensors.5c03166
18. K.-H. Kim et al., "Long-term reliable wireless H_2 gas sensor via repeatable thermal refreshing of palladium nanowire," Nat. Commun. 15 (2024). https://doi.org/10.1038/s41467-024-53080-0
19. W.-T. Koo, J.-S. Jang and I.-D. Kim, "Metal–organic frameworks for chemiresistive sensors," Chem 5 (2019) 1938–1963. https://doi.org/10.1016/j.chempr.2019.04.013
20. N. Thokala et al., "Advancements in porous framework materials for chemiresistive hydrogen sensing: Exploring MOFs and COFs," Dalton Trans. 54 (2025) 3526–3550. https://doi.org/10.1039/D4DT02551B
21. C. Montoro et al., "MOF-derived metal oxide (Cu, Ni, Zn) gas sensors with excellent selectivity towards H_2S, CO and H_2 gases," Compos. Part B Eng. 283 (2024) 111637. https://doi.org/10.1016/j.compositesb.2024.111637
22. V. Martvall et al., "Accelerating plasmonic hydrogen sensors for inert gas environments by transformer-based deep learning," ACS Sens. 10 (2025) 376–386. https://doi.org/10.1021/acssensors.4c02616
23. S. Wang et al., "Deep learning-based hydrogen leakage localization prediction considering sensor layout optimization in hydrogen refueling stations," Process Saf. Environ. Prot. (2024). https://doi.org/10.1016/j.psep.2024.06.122
24. M. Zhou, S. Wang, J. Li, Z. Wei and L. Shui, "A wireless sensor network-based combustible gas detection system using PSO-DBO-optimized BP neural network," Sensors 25 (2025) 3151. https://doi.org/10.3390/s25103151
25. D. Martínez et al., "Experimental application of an autonomous mobile robot for gas leak detection in indoor environments," Proc. 17th Int. Conf. Information Fusion, 2014, pp. 1–7.
26. Z. Zhao, M. Knight, S. Kumar, E. T. Eisenbraun and M. A. Carpenter, "Humidity effects on Pd/Au-based all-optical hydrogen sensors," Sens. Actuators B Chem. 129 (2008) 726–733.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTERACTING MULTIPLE MODEL–BASED ESTIMATION OF AIRSPEED AND ALTITUDE FROM PITOT–STATIC MEASUREMENTS

VLADA SOKOLOVIĆ
Military Academy, University of Defense in Belgrade, Belgrade, vlada.sokolovic@va.mod.gov.rs

MOMIR STANKOVIĆ
Military Academy, University of Defense in Belgrade, Belgrade, momir.stankovic@va.mod.gov.rs

GORAN MARKOVIĆ
School of Electrical Engineering, University of Belgrade, Belgrade, gmarkovic@etf.bg.ac.rs

Abstract: *The accurate estimation of aircraft airspeed and altitude using Pitot–static systems is essential for reliable flight control and navigation. Conventional estimation methods, such as the Extended Kalman Filter (EKF), inherently rely on a single dynamic model and may exhibit degraded performance in the presence of sensor degradation and time-varying measurement disturbances. This motivates the development of estimation algorithms capable of accommodating multiple operating conditions.*

This paper investigates an Interacting Multiple Model (IMM) estimator for air-data parameter reconstruction from the Pitot–static measurements. The estimator combines multiple Extended Kalman Filters, each associated with a specific sensor operating condition, including nominal operation, pressure bias, and sensor drift. A mathematical model of the Pitot–static system is developed, incorporating sensor dynamics, measurement noise, bias, and drift effects. The IMM algorithm is implemented in the MATLAB environment and evaluated under representative aircraft flight scenarios.

The performance of the proposed IMM estimator is evaluated through Monte Carlo simulations and compared with that of a conventional EKF estimator. The obtained results demonstrate that the IMM algorithm achieves improved estimation accuracy in the presence of sensor bias and drift while maintaining reliable estimation of flight parameters under different operating conditions. The obtained results indicate the applicability of IMM-based estimation for the improvement of ADC robustness against sensor degradation, while maintaining reliable air-data parameter estimation.

Keywords: *Air Data Computer, Pitot–static system, Extended Kalman Filter, Interacting Multiple Model, Air-data estimation.*

1. INTRODUCTION

The Pitot–static system represents a critical measurement subsystem in an aircraft, providing air-data information required for flight control, navigation, and other avionics functions. Modern aircraft and unmanned aerial vehicles increasingly employ Micro-Electro-Mechanical Systems (MEMS) sensors to perform pressure and temperature measurements due to their compact size, the low power consumption, and compatibility with digital processing architectures. The measurements of static pressure, total pressure, and outside air temperature are processed by the Air Data Computer (ADC) to determine fundamental flight parameters, including airspeed, altitude, vertical speed, and air density [1], [2].

Regardless of these advantages, MEMS-based air-data measurements remain sensitive to several error sources that directly affect estimation accuracy. The most relevant error sources include measurement noise, temperature dependency, sensor bias, long-term drift, nonlinear sensor characteristics, and dynamic effects which are introduced by the pneumatic response of the Pitot–static installation. Furthermore, environmental influences such as vibration, moisture, and partial blockage of the sensing system may degrade measurement quality and reduce the reliability of the estimated air-data parameters [3].

The Air Data Computer integrates measurements from Pitot–static and temperature sensors by employing digital processing and estimation algorithms in order to provide reliable air-data information for aircraft systems. Previous studies have demonstrated the importance of analytical modeling of Pitot–static sensor faults and the estimation-based approaches for the improvement of fault tolerance and reduction of measurement degradation impact [4]–[6]. Among available estimation techniques, the EKF estimator is widely used in nonlinear state estimation applications because of it's capability to handle nonlinear system models with moderate computational complexity. However, the EKF estimator is generally designed around a predefined system model and may experience degraded performance when measurement characteristics or system behavior change due to sensor degradation or abnormal (i.e. unexpected) operating conditions.

To overcome the limitations associated with single-model estimation, multiple-model approaches have also been investigated for systems that are characterized by varying operating conditions. Among these, the IMM algorithm

Corresponding author: V. Sokolović;
E-mail: vlada.sokolovic@va.mod.gov.rs
DOI: 10.5937/OTEH26056S

represents an established framework for the adaptive state estimation. The IMM algorithm employs several parallel estimators associated with the different system models and updates model probabilities according to consistency between the predicted and the measured outputs, with a goal to facilitate adaptation to changes in system behavior [7]. The theoretical foundation of the IMM approach was introduced by Blom and Bar-Shalom for the systems with Markovian switching parameters, while the subsequent developments extended these multiple-model estimation techniques for variable-structure systems [8], [9].

Previous studies have shown that IMM-based estimators can improve robustness in systems affected by the model uncertainty and changing measurement conditions [10], [11]. Comparative investigations of multiple-model EKF approaches have also shown that estimator performance strongly depends on correct model definition, transition probability selection, and process noise tuning [12].

Although IMM-based algorithms have been extensively investigated in applications such as maneuvering target tracking, navigation, and multi-sensor data fusion, their application to Pitot–static air-data estimation with the explicit representation of sensor bias and drift effects has received considerably less attention. In particular, the development of IMM-based estimators that explicitly incorporate sensor bias and drift models within the air-data measurement process has received less attention. Therefore, this represents an important research direction since these degradation mechanisms may directly affect the accuracy and the reliability of ADC outputs.

The main contribution of this work is the development and evaluation of an IMM-based estimator specifically adapted for Pitot–static air-data processing with explicit representation of pressure sensor bias and drift. The proposed approach employs a bank of Extended Kalman Filters representing different sensor operating conditions, including nominal operation, pressure measurement bias, and sensor drift. A mathematical model of the Pitot–static system is developed that incorporates pneumatic sensor dynamics, measurement noise, pressure measurement bias, and drift characteristics. The proposed estimator is implemented in MATLAB and evaluated through Monte Carlo simulations, while its performance is compared with a conventional EKF estimator. The obtained results provide a methodological basis for a further development of robust Air Data Computer estimation algorithms that are capable to maintain improved accuracy under the degraded sensor conditions.

2. PITOT–STATIC SYSTEM MODEL AND SIMULATION FRAMEWORK

2.1. Mathematical Model of the Pitot–Static System and Air-Data Parameter Computation

A mathematical model of the Pitot–static system was established to generate reference air-data parameters and simulate sensor measurements corresponding to ADC operation. The developed model establishes a relationship between flight conditions, atmospheric parameters, and the outputs of the pressure and temperature sensors. The atmospheric parameters were calculated according to the International Standard Atmosphere model [13]. The air temperature at a given altitude is thus defined as:

$$T = T_0 - Lh, \tag{1}$$

where T_0 represents the standard temperature at sea level, L is the temperature lapse rate, and h is the flight altitude. The static pressure is then calculated using the barometric equation:

$$p_s = p_0 \left(\frac{T}{T_0} \right)^{\frac{g}{RL}}, \tag{2}$$

where p_0 denotes the standard sea-level pressure, R is the specific gas constant of air, and g represents gravitational acceleration. Based on this, the air density is obtained from the ideal gas relationship:

$$\rho = \frac{p_s}{RT_{at}}. \tag{3}$$

The total pressure measured by the Pitot tube is defined as the sum of static and dynamic pressure:

$$p_t = p_s + \rho \frac{V^2}{2}, \tag{4}$$

where V denotes the true airspeed value. The air-data parameters used for estimator evaluation are calculated from the simulated measurements, while the altitude is obtained from the inverse barometric relationship:

$$h = \frac{T_0}{L} \left[1 - \left(\frac{P_s}{p_0} \right)^{\frac{RL}{g}} \right]. \tag{5}$$

The dynamic pressure is determined as difference of total and static pressure, while the airspeed is calculated as:

$$V = \sqrt{\frac{2q}{\rho}}. \tag{6}$$

The vertical speed is obtained by using the numerical differentiation of the estimated altitude:

$$V_z = \frac{dh}{dt}. \tag{6}$$

The calculated air-data parameters are subsequently used as the reference quantities in order to adequately evaluate the performance of the EKF and IMM estimators.

2.2. Sensor Measurement Model and Representation of Operating Modes

The accuracy of a Pitot–static measurement system is influenced by several non-ideal effects, including sensor response delay, measurement noise, systematic offsets,

and long-term variations. In this work, a specific sensor measurement model was developed to represent different operating conditions required for IMM-based estimation.

The dynamic response of the pressure measurement channels was represented by a first-order system:

$$p_{dyn}(k) = p_{dyn}(k-1) + \frac{\Delta t}{\tau}\left(p_{true}(k) - p_{dyn}(k-1)\right), \quad (7)$$

where p_{true} represents the ideal pressure value, p_{dyn} is the dynamically filtered sensor output, τ represents the sensor response time constant, and Δt is the sampling interval.

The measured pressure value is defined as:

$$p_{meas} = p_{dyn} + b + d + n, \quad (8)$$

where b represents sensor bias, d represents sensor drift, and n denotes a measurement noise. The measurement noise is modeled as a zero-mean Gaussian process, and the sensor bias is represented as a slowly varying offset:

$$b(k) = b(k-1) + w_b(k), \quad (9)$$

where $w_b(k)$ represents the bias process noise. The long-term sensor drift is modelled by using a random-walk process:

$$d(k) = b(k-1) + w_d(k), \quad (10)$$

where $w_d(k)$ represents the drift process noise.

The IMM-based models differ not only due to the process noise covariance tuning but also due to the assumed sensor degradation dynamics. To enable IMM operation, three sensor operating modes are defined:

- Model 1 – Nominal operating mode, which represents normal sensor operation, and includes measurement noise and dynamic response effects only;
- Model 2 – Bias mode, which represents the sensor condition that is characterized by a constant pressure measurement offset. In simulation used in this work, the pressure bias of 50 Pa is introduced in the time period between 200 s and 300 s; and
- Model 3 – Drift mode, which is used in this work to represent the gradual sensor degradation after 350 s of simulation time. The drift component is generated using a random-walk process in order to reproduce slowly varying sensor characteristics.

The beforehand defined operating modes provide a physical interpretation of the IMM model set and enable adaptive selection between different sensor conditions.

2.3. Simulation Configuration and Measurement Data Generation

An appropriate MATLAB-based simulation environment was developed to evaluate the proposed IMM estimator. A simulation process includes the flight trajectory model, atmospheric calculations, Pitot–static sensor dynamics, measurement noise, and the sensor degradation scenarios. The total time duration of each simulation is $T_{sim} = 600$ s with the sampling interval defined as $\Delta t = 0.01$ s. The used flight trajectory consists of several climb, level-flight, and descent segments, where altitude and airspeed variations are independently defined according to the selected flight profile.

Based on the generated flight conditions, reference values of altitude, airspeed, temperature, static pressure, and total pressure are calculated. The pneumatic response of the Pitot–static system is represented by first-order dynamics with a different time constants for the pressure channels $\tau_{ps} = 0.1$ s for the static pressure channel and $\tau_{pt} = 0.5$ s for the total pressure channel. The measurement noise is introduced as a Gaussian process with the following standard deviations $\sigma_{ps} = 10$ Pa, $\sigma_{pt} = 5$ Pa and $\sigma_T = 0.5$ K.

The generated sensor measurements are then additionally processed using a second-order Butterworth low-pass filter with a cutoff frequency of 2 Hz to reduce a high-frequency measurement disturbances.

The resulting measurement data are then used as inputs for the both EKF and IMM estimators. Their performance is evaluated under the nominal conditions and during the sensor degradation scenarios that involve bias and drift effects.

The Figure 1 illustrates the simulated and reference Pitot–static sensor outputs used as input data for the evaluation of the EKF and IMM estimators.

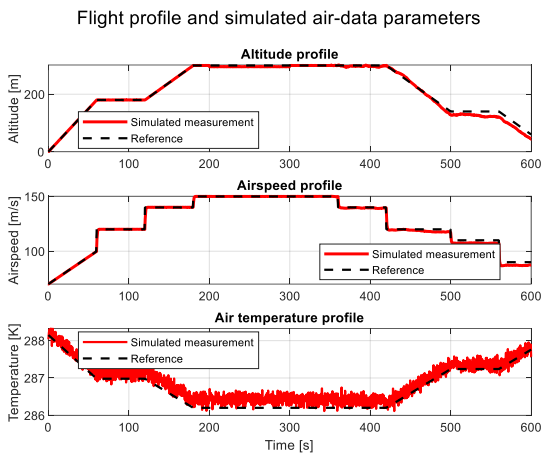


Figure 1. Flight profile and simulated reference air-data parameters used in the Pitot–static system model

3. EXTENDED KALMAN FILTER AND INTERACTING MULTIPLE MODEL ESTIMATION ALGORITHM

3.1. Extended Kalman Filter Formulation for Air-Data Estimation

The Extended Kalman Filter (EKF) is used as a baseline estimator for the evaluation of performance improvement achieved by using the proposed IMM approach. The EKF provides the nonlinear state estimation by linearizing the process and measurement models around the current

operating point. In this work, the estimation is performed directly using Pitot–static pressure measurements, while derived air-data parameters are then obtained from the estimated states. This approach avoids additional filtering of calculated quantities and thus preserves the physical relationships between altitude, airspeed, and atmospheric parameters.

The state vector of the EKF is defined as:

$$\mathbf{x}_k = [h_k \quad V_k \quad T_k \quad b_{p_s,k} \quad b_{p_t,k}], \tag{11}$$

where h_k represents altitude, V_k is airspeed, T_k is air temperature, while b_{ps} and b_{pt} denote the biases of the static and the total pressure sensors, respectively. The nonlinear state transition model is expressed as:

$$\mathbf{x}_k = f(x_{k-1}) + \omega_k, \tag{12}$$

where w_k represents the process noise with covariance matrix **Q**. The flight parameters are assumed to remain constant within a single sampling interval, while the pressure sensor biases are modeled as slowly varying stochastic processes $b_{ps,k} = b_{ps,k}+w_{bps}$ and $b_{pt,k} = b_{pt,k}+w_{bts}$. The measurement vector is defined using the directly available Pitot–static sensor outputs:

$$\mathbf{z}_k = [p_{s,k} \quad p_{t,k} \quad T_k], \tag{13}$$

and the nonlinear measurement equation is given by:

$$\mathbf{z}_k = h(x_k) + \upsilon_k, \tag{14}$$

where v_k represents zero-mean Gaussian measurement noise with covariance matrix **R**.
The EKF prediction step is performed according to:

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= f(\hat{\mathbf{x}}_{k-1}), \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}, \end{aligned} \tag{15}$$

where $\mathbf{F}_k$ is the state transition Jacobian matrix. Since the selected process model assumes that the flight parameters remain constant within the same sampling interval, while sensor biases follow a random-walk process, the state transition matrix is approximated by the identity matrix. The measurement update is calculated using the Kalman gain:

$$\mathbf{K_k} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T \left[\mathbf{H}_k \mathbf{P}_{k-1} \mathbf{H}_k^T + \mathbf{R_k}\right]^{-1} \tag{16}$$

where $\mathbf{H}_k$ represents the measurement Jacobian matrix defined as:

$$\mathbf{H}_k = \begin{bmatrix} \frac{\partial p_s}{\partial h} & 0 & 0 & -1 & 0 \\ \frac{\partial p_t}{\partial h} & \frac{\partial p_t}{\partial V} & \frac{\partial p_t}{\partial T} & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{17}$$

where $\frac{\partial p_t}{\partial h} = \frac{\partial p_s}{\partial h} + \frac{1}{2}V^2 \frac{\partial \rho}{\partial h}$, $\frac{\partial \rho}{\partial h} = \frac{1}{RT}\frac{\partial p_s}{\partial h}$ and $\frac{\partial p_s}{\partial V} = \rho V$, $\frac{\partial p_s}{\partial h} = -\frac{g}{RT} p_s$,

The updated state estimate is obtained as:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - h(\mathbf{x}_{k|k-1})), \tag{18}$$

while the covariance matrix is updated according to:

$$\mathbf{P_k} = (\mathbf{I} - \mathbf{K_k}\mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{19}$$

The EKF estimator described above represents the single-model estimation approach used as a reference for the comparison with the proposed IMM estimator.

3.2. Interacting Multiple Model Estimator Configuration

The IMM estimator is introduced as an extension of the single-model EKF approach described in the previous section. Unlike the conventional EKF, which assumes a fixed system behavior, the IMM algorithm employs a bank of parallel EKF estimators with the different process noise characteristics in order to represent possible (i.e. here observed) sensor operating conditions.
In this work, three beforehand defined operating modes of the Pitot–static system are considered, which correspond to the nominal operation, sensor bias, and sensor drift conditions. In each mode the EKF formulation presented in Section 3.1 is employed, while a different process noise covariance matrices $\mathbf{Q}_1$ $\mathbf{Q}_2$, and $\mathbf{Q}_3$ are assigned in order to adjust the estimator sensitivity to the corresponding sensor behavior.
The IMM model set is therefore defined as:

$$\mathbf{M} = \{M_1 \quad M_2 \quad M_3\}, \tag{20}$$

where each model has an associated EKF estimator and model probability μ_i. The transition between operating modes is described by the Markov transition probability matrix:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}, \tag{21}$$

where π_{ij} represents the probability of transition from model i at the previous time step to model j at the current time step.

At each sampling instant, the IMM estimator performs state interaction, model-conditioned EKF filtering, model probability update, and fusion of individual estimates. In the interaction step, the initial conditions for each EKF are obtained by combining the state estimates and the covariance matrices from all models according to their previous probabilities and the Markov transition matrix. The predicted probability of model j is calculated as:

$$c_j(k) = \sum_{i=1}^{N} \pi_{ij}\mu_i(k-1), \tag{22}$$

where μ_i(k−1) denotes the probability of model i at the previous time step. Corresponding mixing probabilities are determined by:

$$\mu_{ij}(k) = \frac{\pi_{ij}\mu_i(k-1)}{c_j(k)}. \tag{23}$$

These probabilities are used to compute the mixed initial state and covariance for each EKF prior to the prediction and correction steps described in Section 3.1. After the EKF update, the consistency of each model with the current measurements is evaluated using the innovation vector and innovation covariance. The likelihood function of the j-th model is calculated as:

$$\Lambda_j(k) = \frac{1}{\sqrt{(2\pi)^m |S_j(k)|}} \exp(D),$$
$$D = -\frac{1}{2} v_j^T(k) S_j^{-1}(k) v_j(k), \tag{24}$$

where $v_j(k)$ is the innovation vector, $S_j(k)$ is the innovation covariance matrix, and m denotes the dimension of the measurement vector. The model probabilities are then updated according to Bayes' rule:

$$\mu_j(k) = \frac{\Lambda_j(k) c_j(k)}{\sum_{i=1}^{N} \Lambda_i(k) c_i(k)}. \tag{25}$$

Finally, the overall IMM estimate is obtained as the weighted combination of the individual EKF estimates:

$$\hat{\mathbf{x}}_{IMM}(k) = \sum_{i=1}^{N} \mu_j(k) \hat{x}_j(k), \tag{25}$$

while the corresponding covariance matrix is computed as

$$\mathbf{P}_{IMM}(k) = \sum_{i=1}^{N} \mu_j(k) \mathbf{E},$$
$$\mathbf{E} = \mathbf{P}_j(k) + \left(\hat{\mathbf{x}}_j(k) - \hat{\mathbf{x}}_{IMM}(k)\right)\left(\hat{\mathbf{x}}_j(k) - \hat{\mathbf{x}}_{IMM}(k)\right)^T. \tag{26}$$

The resulting model probabilities provide an indication of the most likely sensor operating condition at the each sampling instant, while weighted state estimate combines the advantages of all candidate models. Consequently, the here proposed IMM estimator preserves the estimation accuracy during the nominal operation, and also improves robustness under sensor bias and drift conditions.

The numerical values of the process noise covariance matrices, transition probability matrix, and initial model probabilities used in the simulations are presented in the following section.

4. SIMULATION RESULTS AND DISCUSSION

For the statistical evaluation of the proposed IMM estimator, 100 Monte Carlo simulation runs have been performed under the nominal operating conditions as well as in the presence of sensor bias and sensor drift, using different realizations of process and measurement noise. The simulation parameters adopted throughout this study are summarized in Table 1.

Table 1. Simulation parameters

Parameter	Value
P_0	diag(10^2,10^2,1^2,50^2,50^2)
R	diag(10^2,5^2,0.5^2)
Q_1	diag(0.5^2,0.2^2,0.05^2,0.01^2,0.01^2)
Q_2	diag(0.5^2,0.2^2,0.05^2,1.5^2,1.5^2)
Q_3	diag(0.5^2,0.2^2,0.05^2,0.05^2,0.05^2)
μ_0	$[0.98,0.01,0.01]^T$
Π	[0.995 0.004 0.001; 0.010 0.985 0.005: 0.005 0.010 0.985]

Figure 2 shows that both estimators accurately follow the reference altitude and velocity during nominal operation. However, in the presence of sensor degradation, the IMM estimator achieves better agreement with the reference values compared with the EKF. Unlike the single-model EKF estimator, the IMM estimator adapts the estimation process according to the most probable sensor operating condition, and thus reduce the influence of bias and drift effects.

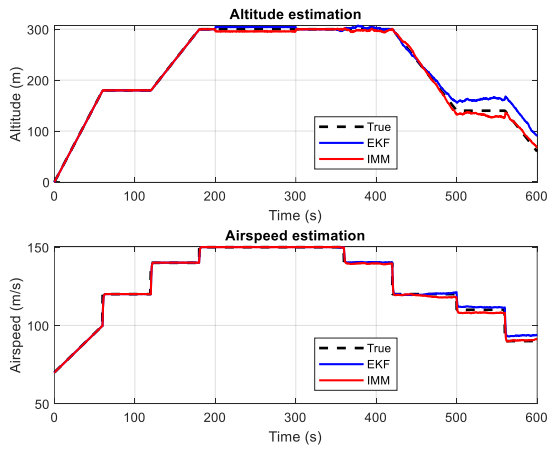


Figure 2. Comparison of EKF and IMM estimates

The Monte Carlo simulation results presented in Figure 3 confirm the robustness of the proposed IMM estimator. For all the analyzed parameters, here proposed IMM approach achieves the lower root mean square error (RMSE) values when compared with the EKF estimator, with the most significant improvements being obtained in case of the altitude estimation.

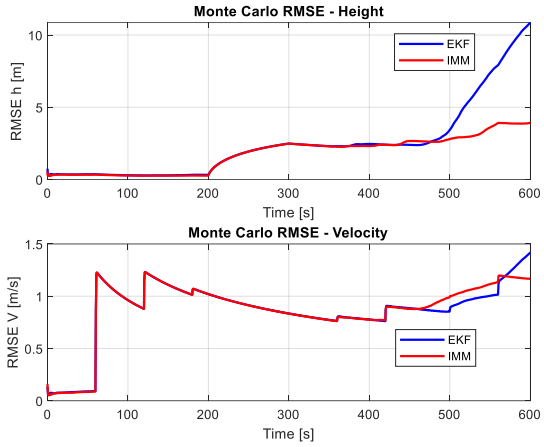


Figure 3. RMSE comparison obtained from Monte Carlo simulations

The quantitative comparison that is presented in Table 2 confirms the advantages of here proposed IMM estimator over the conventional EKF estimator approach. The lower RMSE values are achieved for all analyzed parameters, which indicates the improved estimation accuracy while algorithm stability is also maintained. It shuld be stressed, that the most significant improvement is observed in the case of altitude estimation, which confirms effectiveness of here observed multi-model approach under the sensor degradation conditions.

Table 2. RMSE comparison

State	EKF RMSE	IMM RMSE	Improvement
Altitude (m)	2.4433	1.7859	26.9%
Airspeed (m/s)	0.8669	0.7742	8.3%

5. CONCLUSION

This paper presents an air data parameter estimator based on the IMM algorithm and a bank of EKFs for the different operating regimes of a pitot-static system. The proposed approach includes nominal operation, sensor bias, and drift models, with the adaptive mode selection based on the estimated model probabilities.

The observed nonlinear Pitot–static system model include sensor dynamics, measurement noise, bias, and drift effects that were used for estimator evaluation. The obtained Monte Carlo simulations results show that the IMM approach reduces estimation errors compared with the conventional EKF while preserving stable operation under degraded sensor conditions. Our future work will focus on incorporating additional degradation models and validation of the proposed approach by using real-time experimental flight test data.

Acknowledgments

This work was supported by the Ministry of science, technical development and innovations of Repuplic of Serbia under contract numbe 451 03/200103.

References

[1] HAN, X., HUANG, M., WU, Z., GAO, Y., XIA, Y., YANG, P. et al.: Advances in High-Performance MEMS Pressure Sensors: Design, Fabrication, and Packaging, Microsystems & Nanoengineering, 9(1) (2023).

[2] CHEN, Z., LI, H., YU, H., ZHAO, Y., MA, J., ZHANG, C., ZHANG, H.: Designing of Airspeed Measurement Method for UAVs Based on MEMS Pressure Sensors, Sensors, 24(17) (2024).

[3] PIENIĄŻEK, J., CIECIŃSKI, P., FICEK, D., SZUMSKI, M.: Dynamic Response of the Pitot Tube with Pressure Sensor, Sensors, 23(5) (2023).

[4] LAROCQUE, F., DE PONTI, T., REMES, B., SMEUR, E.: Synthetic Air Data System for Pitot Tube Failure Detection on the Variable Skew Quad Plane, Proceedings of the AIAA SCITECH Forum, 2025.

[5] FREEMAN, P., SEILER, P., BALAS, G. J.: Air Data System Fault Modeling and Detection, Control Engineering Practice, 21(10) (2013) 1290-1301.

[6] KILIC, U., UNAL, G.: Aircraft Air Data System Fault Detection and Reconstruction Scheme Design, Aircraft Engineering and Aerospace Technology, 93(6) (2021) 1104-1114.

[7] RADHIKA, M. N., MALLICK, M., TIAN, X.: IMM Filtering Algorithms for a Highly Maneuvering Fighter Aircraft: An Overview, Algorithms, 17(9) (2024).

[8] BLOM, H. A. P., BAR-SHALOM, Y.: The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients, IEEE Transactions on Automatic Control, 33(8) (1988) 780-783.

[9] LI, X. R., BAR-SHALOM, Y.: Multiple-Model Estimation with Variable Structure, IEEE Transactions on Automatic Control, 41(4) (1996) 478-493.

[10] WANG, S., MEN, C., LI, R., YEO, T. S.: A Maneuvering Extended Target Tracking IMM Algorithm Based on Second-Order Extended Kalman Filter, IEEE Transactions on Instrumentation and Measurement, 73 (2024).

[11] YANG, B., WANG, H., SHI, Z.: Interacting Multiple Model Adaptive Robust Kalman Filter for Process and Measurement Modeling Errors Simultaneously, Signal Processing, 227 (2025).

[12] LI, X. R., JILKOV, V. P.: Survey of Maneuvering Target Tracking. Part V: Multiple-Model Methods, IEEE Transactions on Aerospace and Electronic Systems, 41(4) (2005) 1255-1321.

[13] ISO 2533: Standard Atmosphere, International Organization for Standardization, Geneva, Switzerland, 1975.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SENSOR FUSION FOR DRONE CLASSIFICATION IN RADIO FREQUENCY DOMAIN

DARKO MIKANOVIĆ
Military Technical Institute, Belgrade, darko.mikanovic@mod.gov.rs

BOBAN SAZDIĆ-JOTIĆ
Military Technical Institute, Belgrade, boban.sazdic.jotic@vs.rs

VASILIJA JOKSIMOVIĆ
Military Technical Institute, Belgrade, vasilija.joksimovic@mod.gov.rs

Abstract: *The increasing usage of Unmanned Aerial Systems in critical areas requires robust Counter Unmanned Aerial Systems solutions to maintain security and operational integrity. This paper investigates the impact of sequential sensor data fusion on the classification accuracy of Unmanned Aerial Systems signals within the Radio Frequency domain for real-time inference. Given the hardware constraint of a single Radio Frequency receiver, we propose a tailored fusion architecture that integrates sequential data acquired over time. Utilizing Radio Frequency spectrograms as the primary input, various sensor fusion methods were evaluated in combination with simple Convolutional Neural Networks. To ensure compatibility with our deep learning models, particularly under the constraints of Radio Frequency sensors that are often affected by low Signal-to-Noise Ratio and spectral interference in dense environments, the proposed fusion methods produce post-processed spectrograms that retain the original input dimensions. This design enables seamless integration into pre-trained deep learning pipelines without necessitating architectural modifications, while also addressing conditions that may otherwise contribute to elevated false-positive rates. The proposed methods were benchmarked against multiple pre-recorded validation datasets specifically designed to simulate real-world scenarios. The results demonstrate that specific temporal sensor fusion approaches significantly enhance prediction accuracy and overall decision-making robustness, providing a more stable decision boundary compared to baseline models without data fusion. Ultimately, the proposed approach offers a computationally efficient technique to improve inference reliability without additional overhead hardware or extensive model fine-tuning. By effectively mitigating stochastic noise and signal fading, this method reduces false alarm rates by up to 15%, suggesting that incorporating time-domain sensor fusion as a preprocessing layer is a scalable, cost-effective solution for real-time drone classification in complex Radio Frequency environments.*

Keywords: *classification, drone, convolutional neural network, fusion, radio frequency, sensors.*

1. INTRODUCTION

The development of Unmanned Aerial Systems (UAS) or colloquial “drones” and the technology they use is progressing at an incredible pace, while the reduction in their price and the simplification of their use contribute to their increasing availability and accessibility. As a result, there has been a need for protection against malicious drones that seek to damage critical structures in some way. This is how Counter Unmanned Aerial Systems (CUAS) emerged, which, with the help of sensor systems and Artificial Intelligence (AI) algorithms, detect and identify drones and react appropriately in order to neutralize the threat [1].

Most AI algorithms developed for this application rely primarily on individual sensor inputs, such as images, whereas real-time scenarios in which sensors continuously generate large volumes of image data are typically not sufficiently addressed [2]. When used in real-time, detection algorithms often give too many decisions per unit of time, and the end-user does not benefit from that. Moreover, such decisions are imperfect because of the detection algorithms themselves, which have a certain accuracy over trained datasets that are generally not large and diverse enough due to unavailability of adequate training data and the very nature of the problem. This is strictly connected with the characteristics of Radio Frequency (RF) environment often affected by low Signal-to-Noise Ratio (SNR), spectral interference and propagation of electromagnetic waves [3], [4], [5].

The research subject includes the introduction of automatic data fusion into a system with a single sensor for detection and identification of drones in the RF domain. Multiple early and late fusion techniques were tested through experimental research (spectrogram merging, hard and soft fusion), and the corresponding results are presented offering the tailored approach for this specific problem. Each of the fusion approaches are explained in the Methodology chapter of the paper. The main goal is to increase the reliability and accuracy of drone detection and identification CUAS systems that

Corresponding author: D. Mikanović;
E-mail: darko.mikanovic@mod.gov.rs
DOI: 10.5937/OTEH26057M

leverage already trained Convolutional Neural Network (CNN) models for predictions.

There are several advantages of using optimal sensor fusion techniques. The reliability and accuracy of drone detection and identification are increased with very little overhead and using the available prediction models as-is. The number of decisions made per unit of time is brought closer to the optimum for the end-user. An optimal sensor fusion is also a good starting point for more advanced data fusion approaches that include other types of fusion, such as post-prediction late fusion, that can increase the accuracy and reliability of detection systems even more.

The limitation of the sensor fusion approach is that it cannot improve the predictions accuracy by much on its own without other data fusion layers, though incorporating sensor fusion along with advanced feature fusion or late fusion methods tend to bring the best results.

The main contribution of this work consists of several facts: improvement of the classification accuracy in CNN models with temporal data fusion approaches, examination of the classification accuracy in 2,4GHz frequency band, testing the possibility of detecting multiple drones operating at the same time, and novel benchmark for examining the classification accuracy of drone classification in RF domain.

The rest of the paper is organized as follows: chapter Literature presents the recent research in this area specifically solving this problem, chapter Methodology presents the system model and explains to the reader how the experiments are performed, chapter Results presents the obtained results with discussion, and finally chapter Conclusion concludes the study.

2. LITERATURE

2.1 RF Domain and Spectrogram Analysis for UAS Classification

The rapid proliferation of UAS within critical airspace underscores the immediate necessity for dependable CUAS technologies [6]. Traditional drone monitoring techniques—primarily dependent on visual recognition, electro-optical sensors (cameras), and acoustic arrays—exhibit severe performance drop-offs when subjected to high ambient noise, physical occlusions, and adverse weather conditions [4], [7], [8]. To bypass these vulnerabilities, RF sensing has emerged as an indispensable modality, offering an all-weather, long-range, non-line-of-sight strategy by intercepting telemetry and control packets exchanged between a drone and its ground control station [9]. Recent trends favor Deep Learning (DL) and AI pipelines over hand-crafted statistical extractions to interpret these signal environments [7], [10], [11]. Specifically, by applying the Short-Time Fourier Transform (STFT), researchers transform one-dimensional raw time-series data or In-phase and Quadrature (I/Q) samples into two-dimensional visual spectrogram layouts [3], [5], [12], [13]. This processing stabilizes signal variance, preserves underlying temporal-spectral fluctuations, and maps the physical channel behaviors directly into joint time-frequency domains optimal for CNN feature backbones [3], [5], [11].

2.2 Multi-Sensor and Cross-Modality Data Fusion Architecture

While 2D DL models yield high accuracy profiles under pristine configurations, real-world RF landscapes remain prone to multipath scattering, intermittent fading, and deliberate or stochastic spectral noise, inflating false-positive metrics [3]. To reinforce system robustness, cross-modality data fusion frameworks have been widely engineered. For instance, multi-sensory CUAS architecture combining heterogeneous sensors—such as radars, electro-optical sensors (cameras), acoustic arrays, and RF sensors—leverage complementary physical properties to yield substantial improvements in drone verification accuracy, spatial classification, and real-time tracking efficiency [7], [10], [14], [15], [16]. Within specific dual-stream frameworks, DL models have been successfully fused with parallel convolutional channels to combine raw acoustic waveforms with RF signatures, illustrating high tracking efficacy across heterogeneous clusters [4], [9]. Furthermore, multi-modal network schemes are frequently supported by advanced algorithmic backbones, such as contrastive learning Siamese topologies or Kalman filters, to reliably identify unauthorized payloads and stealth drone threats under highly complex electromagnetic interference [3], [11], [17].

2.3 Single RF sensor Preprocessing Paradigms and the Temporal RF Fusion Gap

Despite the high detection performance guaranteed by multi-sensor deployment, the physical implementation of distributed, multi-modal hardware remains heavily constrained. Deploying physical radar arrays alongside synchronized electro-optical sensors (cameras) and acoustic arrays introduces severe system configuration complexity, excessive computational workloads, and aggressive calibration costs [9]. Consequently, modern methodologies are actively pivoted toward data reconstruction and virtual feature fusion frameworks capable of extracting hidden or multi-feature temporal streams inside a single-sensor hardware pipeline [12]. Synthesizing sequential temporal segments or multi-scale data profiles prior to or within intermediate convolutional layers provides a highly stable decision boundary, maximizing the model's macro-precision and robustness without scaling hardware boundaries [5], [12].

However, to the best of our knowledge, analysis of existing CUAS literature reveals a stark architectural dichotomy. Current state-of-the-art data fusion solutions are overwhelmingly dependent on either multi-modal hardware configurations (fusing RF with radar [1], [4], [7], [8], [10], [14], [15], [16], [17], vision [9], [18], or acoustics [4]) or concurrent multi-channel array nodes [3]. There remains a severe scarcity of scientific research evaluating sequential, time-domain data fusion

architectures engineered specifically for single-receiver RF infrastructures. Most single-sensor designs still rely on isolated, single-frame spectrogram inferences that lack temporal context and fail during deep signal fading or transient interference bursts. This critical research gap establishes a clear demand for lightweight, computationally efficient preprocessing frameworks capable of dynamically aggregating temporal RF states into standardized dimensions, optimizing classification accuracy without demanding overhead hardware or resource-intensive model fine-tuning.

3. METHODOLOGY

3.1 Fusion Architecture

The proposed system (Figure 1) processes RF spectrograms produced by a single receiver through two independently pre-trained CNN classifiers, one for counting drones (Model 1) and the other for identifying a single drone (Model 2). These models were also used in paper [13], where they are described in detail. Model 1's classes are: 0 drones, 1 drone, 2 drones and 3 drones; model 2's classes are: Lightbridge (0), OcuSync20 (1) and OcuSync40 (2). Model 2's predictions are relevant only when model 1 outputs 1 drone as the prediction. Combining both models' predictions, the proposed data fusion approach has 6 possible decisions to make:

- No drones (model 1: 0),
- Lightbridge (model 1: 1, model 2: 0),
- OcuSync20 (model 1: 1, model 2: 1),
- OcuSync40 (model 1: 1, model 2: 2),
- Two drones (model 1: 2),
- Three drones (model 1: 3).

Spectrograms are recorded in advance and grouped into four different scenarios which are used in this research for fusion evaluation and they will be explained in the Experimental setup section and presented in the Results and discussion chapter. Two families of data fusion are considered: early (sensor) fusion, which merges several successive spectrograms into a single input prior to classification, and late (hard and soft) fusion, which combines the classifiers' outputs over several consecutive predictions. Sensor and late fusion can also be combined, and every technique is configurable in terms of how many spectrograms or predictions are aggregated. Each fusion technique will be described in the following sections.

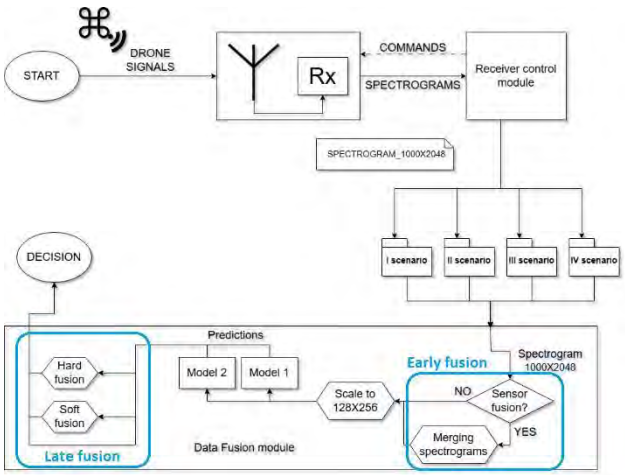


Figure 1. Methodology

3.2 Sensor (Early) Fusion

Since only a single RF sensor is available, sensor fusion here is limited to merging successive spectrograms from that sensor into one spectrogram matching the models' input size. Three merging methods were implemented: vertical concatenation of successive spectrograms followed by rescaling, which compresses drone signal patterns along the time axis while preserving their width — the primary feature distinguishing drone types (Figure 2); pixel-wise logical OR of the aligned spectrograms, which preserves the original dimensions but accumulates noise and interference from every merged frame (Figure 3); and pixel-wise logical XOR, intended to reduce this noise accumulation relative to OR at the cost of a more fragmented resulting signal (Figure 4). Vertical concatenation showed the best results among these three methods in qualitative tests in the 2,4GHz frequency band use case, where the accuracy is increased only slightly when compared to late fusion approaches.

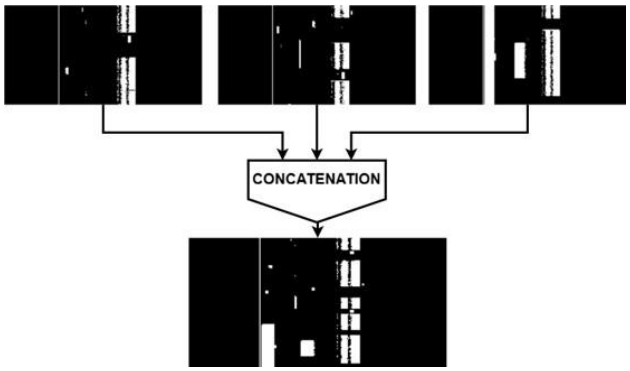


Figure 2. Sensor Fusion – Concatenation

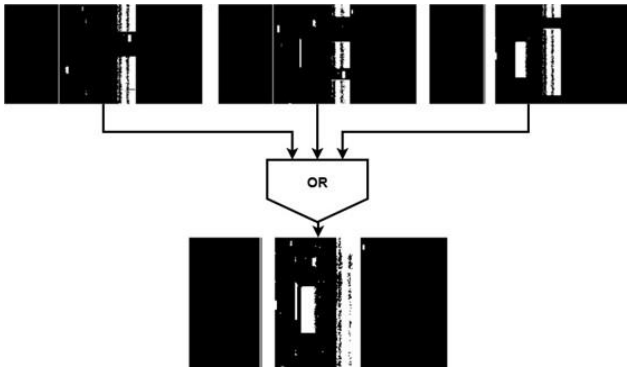


Figure 3. Sensor Fusion – OR operation

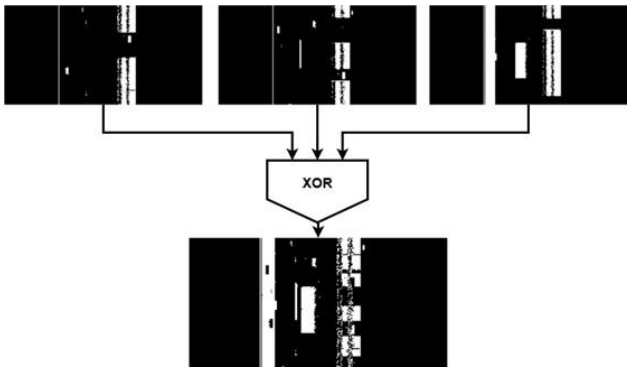


Figure 4. Sensor Fusion – XOR operation

3.3 Score-Level Fusion: Soft Fusion

Soft fusion combines the probability vectors produced by a classifier over several consecutive spectrograms by summing them element-wise and selecting the class with the highest cumulative probability. The soft fusion additionally folds the previous decision window's summed vector into the current sum, scaled by an empirically obtained factor k = 0.25. This makes fusion outcomes more resistant to short-lived state changes, at the cost of slightly slower adaptation in environments where drones frequently appear or disappear, and is therefore best suited to relatively stable operating environments.

3.4 Decision-Level Fusion: Hard Fusion

Hard fusion instead combines the discrete class predictions made over several consecutive spectrograms, using at least three predictions to avoid frequent ties. Under the majority voting rule, the final decision is the class occurring most often within the prediction window (e.g., for predictions [0, 1, 0, 2, 2, 0], the majority voting outcome is class 0). Another rule that is implemented is the longest sequence rule, where the final fusion decision is the class that makes the longest consecutive sequence in the prediction window (e.g., for predictions [0, 0, 1, 1, 1, 0, 0], the longest sequence outcome is class 1). Residual ties are resolved through a configurable secondary rule.

3.5 Experimental Setup

To emulate realistic operating conditions, all recordings were made indoors, where sporadic 2,4GHz interference from other wireless devices is present, and drone signals were intentionally attenuated by placing the drones at a distance from the receiving antenna. Manual band switching between 2,4GHz and 5,8GHz and drone activation/deactivation were used to simulate dynamic scenarios. Four scenarios were each recorded once, so that every fusion configuration is evaluated on identical data: (I) a single Lightbridge drone featured in 108 spectrograms; (II) a single OcuSync40 drone featured in 72 spectrograms; (III) two drones (Lightbridge and OcuSync20) in 108 spectrograms; and (IV) three drones (Lightbridge, OcuSync20 and OcuSync40), 252 spectrograms – each scenario including drone-free intervals. Sensor fusion was limited to merging at most four spectrograms, and late fusion windows were capped at nine consecutive predictions, reflecting the real-time constraint of at most one fusion decision every three seconds at an acquisition rate of three spectrograms per second.

4. RESULTS AND DISCUSSION

Table 1 summarizes, for each scenario, the classification accuracy obtained without fusion and the best-performing early and late fusion configurations identified during testing. Fusion decisions are compared against ground truth at the timestep granularity determined by the corresponding fusion window. Early fusion results in scenarios II and III are omitted because no early fusion configuration could achieve the accuracy of models without fusion due to scenarios' nature. N_{sp} is the number of successive spectrograms used for sensor fusion.

Table 1. Classification accuracy by scenario

Scenario	No fusion	Best early fusion	Best late fusion
I	85.19%	89.81% ($N_{sp} = 2$)	**98.15%** (hard, 4 predictions)
II	91.67%	–	**97.22%** (soft, 3 predictions)
III	96.30%	–	96.30% (soft, 3 predictions)
IV	80.95%	84.52% ($N_{sp} = 3$)	**89.29%** (hard, 4 predictions)

The used scenarios and relevant results are illustrated using plots in the upcoming figures. The blue strip is used to mark the ground truth, that is, the true state at the given time in the scenario. Red dots represent models' predictions without any fusion influence. Models' predictions tend to fluctuate frequently when no fusion method is used to compensate. Finally, the golden stars represent fusion's decisions. The stars appear less frequently than the red dots, since fusion algorithms take several predictions from model's to make a single decision. Number of taken predictions is a configurable parameter.

Scenario I, characterized by pronounced interference, benefited most from fusion: sensor fusion via concatenation of two successive spectrograms raised accuracy from 85.19% to 89.81% (Figure 5), while decision-level hard fusion with four predictions using the majority voting rule reached 98.15% (Figure 6).

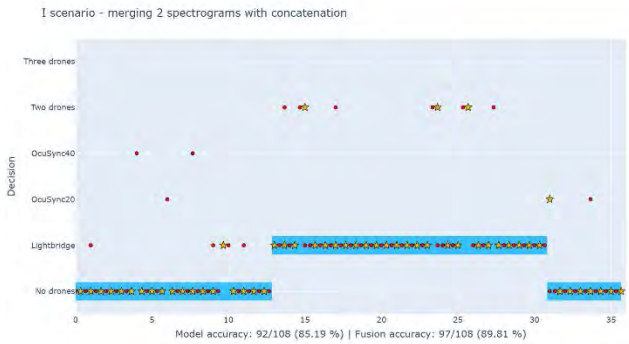


Figure 5. Scenario I – Early fusion

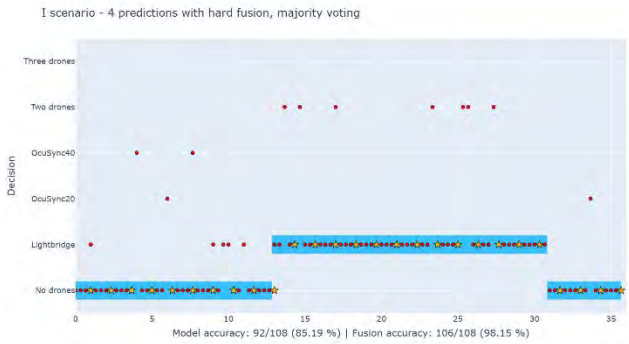


Figure 6. Scenario I – Hard fusion

In the cleaner scenarios II (Figure 7) and III (Figure 8), sensor fusion could not improve the unfused models' accuracy. Soft fusion over three predictions matched or exceeded the unfused baseline in these scenarios (97.22% and 96.30%, respectively) while producing visibly more stable decision sequences than the individual model predictions, which fluctuated at several points despite comparable aggregate accuracy.

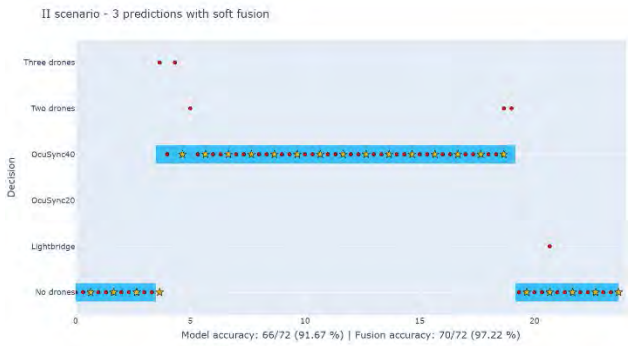


Figure 7. Scenario II – Soft fusion

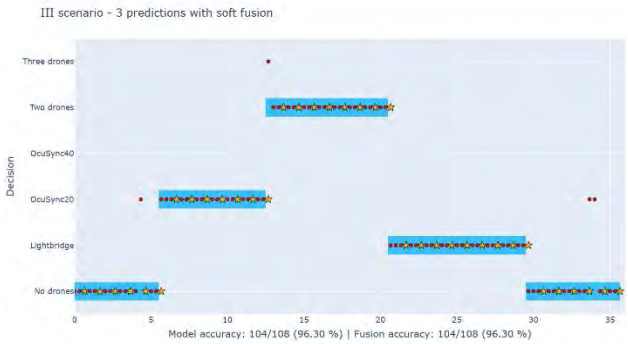


Figure 8. Scenario III – Soft fusion

In the most complex scenario IV (Figures 9 and 10), involving three drones and frequent state changes, hard fusion over four predictions using the longest sequence rule improved accuracy from 80.95% to 89.29%.

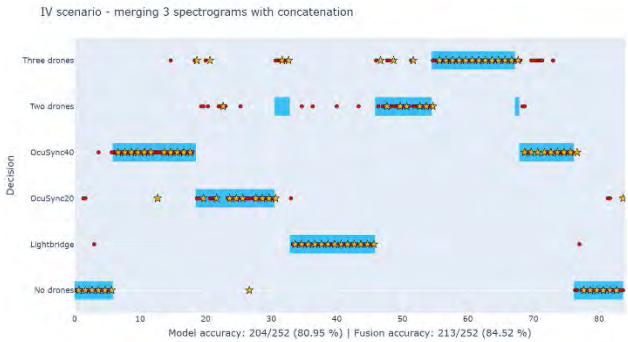


Figure 9. Scenario IV – Early fusion

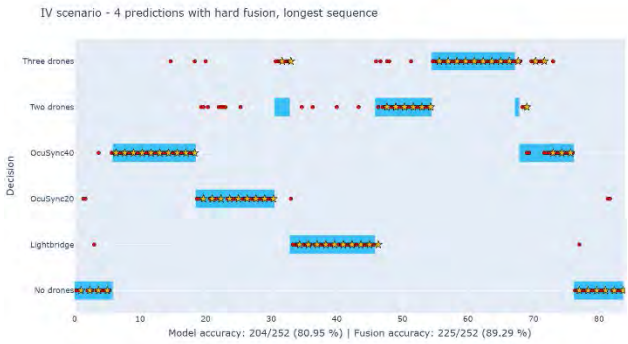


Figure 10. Scenario IV – Hard fusion

Across all scenarios, late fusion consistently outperformed sensor fusion alone, confirming that pairing sensor fusion with a late fusion stage, rather than relying on sensor fusion in isolation, is necessary to obtain the largest accuracy gains. Models used should also have certain amount of accuracy so the fusion algorithms can work out a reliable decision using model's predictions as a baseline.

5. CONCLUSION

This work evaluated sensor and late data fusion for single-receiver RF drone classification with CNN models across four realistic multi-drone scenarios. Sensor fusion alone provided only marginal accuracy gains in the 2,4GHz frequency band. Late fusion, in contrast, improved classification accuracy by up to 13% over the

unfused baseline and produced markedly more stable decisions over time, while also allowing the decision-making frequency to be tuned to the target application.

These findings support time-domain fusion as a lightweight, hardware-free preprocessing layer for improving the precision and reliability of real-time RF-based drone classification. As a result, the predictions that are shown to users in real time became stable and more useful, without fluctuations and excessive output. Future work includes extending the sensor system with additional RF or complementary sensors, fine-tuning the CNN models directly on fused spectrograms, and evaluating further soft and hard fusion strategies.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-34/2026-03/200325).

References

[1] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-drone system with multiple surveillance technologies: architecture, implementation, and challenges," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 68–74, 2018, doi: 10.1109/MCOM.2018.1700430.

[2] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the Detection of Unauthorized Drones—Techniques and Future Perspectives: A Review," *IEEE Sens. J.*, vol. 22, no. 12, pp. 11439–11455, Jun. 2022, doi: 10.1109/JSEN.2022.3171293.

[3] H. Zhang, T. Li, Y. Li, J. Li, O. A. Dobre, and Z. Wen, "RF-Based Drone Classification Under Complex Electromagnetic Environments Using Deep Learning," *IEEE Sens. J.*, vol. 23, no. 6, pp. 6099–6108, Mar. 2023, doi: 10.1109/JSEN.2023.3242985.

[4] A. Frid, Y. Ben-Shimol, E. Manor, and S. Greenberg, "Drones Detection Using a Fusion of RF and Acoustic Features and Deep Neural Networks," *Sensors*, vol. 24, no. 8, p. 2427, Apr. 2024, doi: 10.3390/s24082427.

[5] Y. Fu and Z. He, "Radio Frequency Signal-Based Drone Classification with Frequency Domain Gramian Angular Field and Convolutional Neural Network," *Drones*, vol. 8, no. 9, p. 511, Sep. 2024, doi: 10.3390/drones8090511.

[6] B. Sazdić-Jotić, I. Pokrajac, J. Bajčetić, and N. Stefanović, "Review of RF-based drone classification: Techniques, datasets, and challenges," *Vojnotehnicki glasnik*, vol. 72, no. 2, pp. 764–789, Apr. 2024, doi: 10.5937/vojtehg72-49286.

[7] V. Semenyuk, I. Kurmashev, A. Lupidi, D. Alyoshin, L. Kurmasheva, and A. Cantelli-Forti, "Advances in UAV detection: integrating multi-sensor systems and AI for enhanced accuracy and efficiency," *International Journal of Critical Infrastructure Protection*, vol. 49, p. 100744, Jul. 2025, doi: 10.1016/j.ijcip.2025.100744.

[8] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," *Sensors*, vol. 24, no. 1, p. 125, Dec. 2023, doi: 10.3390/s24010125.

[9] H. Lee *et al.*, "CNN-Based UAV Detection and Classification Using Sensor Fusion," *IEEE Access*, vol. 11, pp. 68791–68808, 2023, doi: 10.1109/ACCESS.2023.3293124.

[10] S. Verma, O. P. Verma, H. Gupta, T. K. Sharma, and S. Agarwal, "AI-Driven Efficient and Reliable Anti-UAV Monitoring System Using Sensor Data Fusion: A Short Report," in *Lecture Notes in Networks and Systems*, vol. 1343 LNNS, Springer Science and Business Media Deutschland GmbH, 2025, pp. 613–623. doi: 10.1007/978-981-96-5955-5_52.

[11] T. Li, Z. Hong, Q. Cai, L. Yu, Z. Wen, and R. Yang, "BisSiam: Bispectrum Siamese Network Based Contrastive Learning for UAV Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12109–12124, Dec. 2023, doi: 10.1109/TKDE.2021.3118727.

[12] J. E. Ko, S. Kim, J. H. Sul, and S. M. Kim, "Data Reconstruction Methods in Multi-Feature Fusion CNN Model for Enhanced Human Activity Recognition," *Sensors*, vol. 25, no. 4, p. 1184, Feb. 2025, doi: 10.3390/s25041184.

[13] B. Sazdić-Jotić, I. Pokrajac, J. Bajčetić, B. Bondžulić, and D. Obradović, "Single and multiple drones detection and identification using RF based deep learning algorithm," *Expert Syst. Appl.*, vol. 187, Jan. 2022, doi: 10.1016/j.eswa.2021.115928.

[14] C. A. Kumar and O. Giridhar, "UAV Detection Multi-sensor Data Fusion," *Journal of Research in Science and Engineering*, vol. 6, no. 7, pp. 6–12, Jul. 2024, doi: 10.53469/jrse.2024.06(07).02.

[15] M. Aledhari, R. Razzak, R. M. Parizi, and G. Srivastava, "Sensor Fusion for Drone Detection," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, IEEE, Apr. 2021, pp. 1–7. doi: 10.1109/VTC2021-Spring51267.2021.9448699.

[16] J. Dudczyk, R. Czyba, and K. Skrzypczyk, "Multi-Sensory Data Fusion in Terms of UAV Detection in 3D Space," *Sensors*, vol. 22, no. 12, p. 4323, Jun. 2022, doi: 10.3390/s22124323.

[17] V. Mehta, F. Dadboud, M. Bolic, and I. Mantegh, "A Deep Learning Approach for Drone Detection and Classification Using Radar and Camera Sensor Fusion," in *2023 IEEE Sensors Applications Symposium (SAS)*, IEEE, Jul. 2023, pp. 01–06. doi: 10.1109/SAS58821.2023.10254123.

[18] Y. Wan, H. Liang, G. Wu, F. Zhou, B. Crispo, and Q. Wu, "Precise RF-Vision Fusion UAV Positioning and Identification for 6G Spectrum Security," *IEEE Journal on Selected Areas in Communications*, vol. 44, pp. 3949–3964, 2026, doi: 10.1109/JSAC.2026.3671229.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

FEDERATED LEARNING FOR AUTONOMOUS DISTRIBUTED GNSS JAMMING SWARMING STATIONS

BOBAN SAZDIĆ-JOTIĆ
Military Technical Institute, Belgrade, boban.sazdic.jotic@vs.rs

DRAGANA BOJIĆ
Military Technical Institute, Belgrade, dragana.bojic@mod.gov.rs

RADIANA BRUSIN
Military Technical Institute, Belgrade, radiana.brusin@mod.gov.rs

Abstract: *The wide implementation of Global Navigation Satellite Systems has declared these technologies critical assets in both military operations and civilian infrastructure. Consequently, Global Navigation Satellite Systems have become a critical target for Electromagnetic Spectrum Operations. Traditional centralized jamming systems, however, are often inadequate in complex and dynamic electromagnetic environments and in large geographic areas due to weaknesses such as single-point failures, increased latency, and restricted signal range. To overcome these limitations, there is a need for autonomous, distributed jamming architectures capable of adapting to dynamic electromagnetic environments. This paper proposes a decentralized approach powered by Federated Learning to enable a swarm of jamming agents to collaboratively learn and execute electronic attack. Federated Learning was selected to reduce communication overhead and enable effective coordination among multiple agents in complex, large-scale environments where synchronization is essential for neutralizing dynamic military targets. To evaluate system performance, we observed the Jamming-to-Signal Ratio at the Global Navigation Satellite Systems receiver point, ensuring the effectiveness of the distributed electronic attack strategy. Our results show that this approach provides excellent adaptability in non-stationary, high-mobility scenario compared to traditional centralized jamming strategies. Exploiting the collective intelligence of multiple autonomous agents and improving the Jamming-to-Signal Ratio at the target receiver, the proposed approach achieves higher operational efficiency and resilience against counter-jamming measures. The Federated Learning framework sets up a scalable and robust foundation for future multi-agent coordination in military electromagnetic spectrum operations.*

Keywords: *agents, artificial intelligence (AI), electromagnetic spectrum operations (EMSO), Federated Learning (FL), Global Navigation Satellite Systems (GNSS), Jamming-to-Signal Ratio (JSR).*

1. INTRODUCTION

Electromagnetic Spectrum Operations (EMSO) include coordinated combat military actions to profile, exploit, and protect the electromagnetic environment, combining electronic warfare (EW) with spectrum management. The primary purpose of EMSO is to achieve superiority and unrestricted access to the electromagnetic spectrum (EMS) for friendly forces while at the same time denying its use to foes. In modern multi-domain conflicts, this is obtained by enabling precision electronic attack (EA), covered and secure communication, and early warning capabilities. However, modern EMSO cope with severe limitations due to the extreme congestion of the radio frequency (RF) environment, the proliferation of commercial off-the-shelf (COTS) software-defined radios (SDR), and the vulnerability of centralized EW architectures [1].

Crucial and specific aspect of successful EMSO is the exploitation of possibly weakens point which are Position, Navigation, and Timing (PNT) signals, most notably the Global Navigation Satellite System (GNSS) [2]. The primary military purpose of GNSS, that engages satellite constellations such as GPS, GLONASS, Galileo, and BeiDou, is to serve as the primary and global tool for giving precise, real-time spatial and temporal information. The critical limitation of GNSS comes from its extreme vulnerability to electronic interference [3], [4], both intentional and unintentional. Unintentional interference arises from civilian infrastructure, space weather phenomena, or RF cosmic radiation and harmonic emissions from nearby communication equipment, which inadvertently degrade sensitivity of GNSS receiver (GNSS RX). Intentional interference stands for a deliberate tactical deployment and engagement of EA assets. From an EMSO perspective, GNSS jamming and spoofing offers profound tactical benefits by creating localized denial-of-service (DoS) zones for PNT signals [4].

In response to these general threats, modern defensive doctrines have integrated advanced GNSS protection mechanisms, complicating the EMSO scene [5]. Current military GNSS RXs are increasingly equipped with anti-jamming capabilities, such as spatial-temporal anti-jamming processing (STAP) [6], Controlled Reception Pattern Antennas (CRPA) [7], and multi-constellation fallback algorithms. While these electronic protection (EP) techniques enhance PTN resilience, they turn the EMS into

Corresponding author: B. Sazdić-Jotić;
E-mail: boban.sazdic.jotic@vs.rs
DOI: 10.5937/OTEH26058S

a highly dynamic, asymmetric environment. For offensive EMSO, traditional high-power, centralized jamming stations or jammer (JX) are no longer fully effective, as protection techniques can nullify static interference sources [7].

To overcome these sophisticated defense techniques, this paper proposes a distributed autonomous GNSS jamming framework based on artificial intelligence (AI). Rather than relying on a single vulnerable JX, the proposed architecture coordinates a swarm of geographically distributed, low-cost and low-power JX capable of executing synchronized jamming [1], [8]. To navigate dynamic RF environments, each JX employs Reinforcement Learning (RL) to autonomously adapt its position and power based on real-time feedback. To address communication limits and data privacy constraints, a Federated Learning (FL) framework [9] is integrated across the swarm, allowing jammers to collaboratively optimize a global model by exchanging compact weight updates instead of raw telemetry [10], [11], [12].

The integration of RL and FL for distributed GNSS jamming offers key EMSO advantages:

- **Resilience.** Spatial distribution eliminates single points of failure, protecting against hard-kill threats.
- **Adaptability.** RL-driven local control allows real-time optimization of position and power against dynamic electronic protection (EP) upgrades.
- **Anti-Jam Evasion.** Multiple low-power JXs effectively bypass modern protections like Controlled Reception Pattern Antennas (CRPA).
- **Communication Efficiency.** FL preserves bandwidth and privacy by exchanging compact model weights instead of raw telemetry.

Conversely, architecture faces operational challenges, including swarm convergence complexity, heterogeneity penalties during federated aggregation (due to varying hardware/communication limits), and the requirement for strict synchronization between JXs to prevent mutual self-interference.

The rest of this paper is organized as follows: chapter 2 presents the literature review, chapter 3 shows the architectural design of the proposed approach, chapter 4 shows the results and discussion, and chapter 5 summarizes the core findings of the study.

2. LITERATURE

2.1. GNSS OVERVIEW

Technologies of GNSS have become a crucial part of modern infrastructure, with broad applications in transportation, telecommunications, scientific research, and numerous civilian and military systems. In recent years, GNSS is again in the focus because of the recent military conflicts around the world. Official GPS interface specification documents [13] provide the basis for understanding the structure of PTN signals, message formats, and the characteristics of individual frequency bands. Additionally, authors in [14] addressed GNSS modernization and analyzed the development of new signals and services, as well as improvements in positioning accuracy. New ideas are presented in [15] giving attention to multi-constellation GNSS RXs, which provide greater positioning reliability and availability in complex environments. Finally, authors in [16] investigated the application of AI to improve GNSS performance, particularly in position estimation and signal-quality analysis.

2.2. GNSS INTERFERENCE (JAMMING & SPOOFING)

Signals of GNSS reach the Earth's surface at extremely low power levels, generally below the GNSS RX thermal-noise floor, which makes it particularly vulnerable to various forms of interference and deliberate attacks. Jamming [17] and spoofing [5] are widely regarded as the most serious EA threats to reliable GNSS operation. In addition, review studies in [18], [19], analyze the effects of interference on GNSS RX performance using parameters such as automatic gain control (AGC), carrier-to-noise density ratio (C/N0), signal spectral characteristics, and in-phase/quadrature (I/Q) samples. Additionally, authors in [20] introduced distributed systems for detecting and localizing interference sources, as well as to sensor networks for monitoring large geographical areas. Modern approaches in [21] increasingly employ AI algorithms for the automatic detection and classification of different interference types. Although many existing algorithms achieve high detection accuracy, most rely on centralized data processing, which represents a limitation in distributed GNSS systems.

2.3. REINFORCEMENT LEARNING AND FEDERATED LEARNING

The integration of RL approaches in the domain of GNSS technologies represents a fundamental innovation in modern EMSO, particularly under conditions of intentional jamming and spectrum vulnerability [22], [23]. By interacting with dynamic EMS, intelligent autonomous unmanned systems can adopt alternative navigation strategies and mitigate GNSS loss [24]. This capability enables systems to adapt to unpredictable spatial variations without requiring predefined, static scenarios. Moreover, FL is a modern distributed AI approach that enables collaborative model training without exchanging raw data among agents. In communication networks, edge computing, and distributed systems, FL has been successfully applied to distributed spectrum sensing, interference classification, and jammer recognition [25]. In the GNSS domain, recent studies demonstrate the potential of FL for spoofing-attack detection [26], GNSS interference classification [27], and the development of personalized models for different receivers and operating environments.

Despite substantial progress, the application of RL and FL in EMSO for distributed GNSS jamming remains underexplored, especially in scenarios involving autonomous JXs and real-time data processing. This research gap provides the primary motivation for the solution proposed in this paper.

3. METHODOLOGY

This chapter defines the proposed simulation model for analyzing and improving the AI distributed jamming of GNSS RX in the context of EMSO. The system model combines RL algorithm for the autonomous positioning of *N* JXs organized in a swarm with a simple FL framework to collaboratively watch and evaluate the cooperative EA effects. The simulation model defines a two-dimensional square, Region of Interest (ROI), $\mathcal{A}$, centered at the coordinate origin, and defined by the set:

$$\mathcal{A} = \{(x, y) \in \mathbb{R}^2 | -D/2 \le x \le D/2, -D/2 \le y \le D/2 \quad (1)$$

where *x* and *y* are planar coordinates, and $D \times D$=20×20 km are ROI dimensions (empirically chosen for operational EMSO level).

3.1. ROI CONFIGURATION AND RASTERIZATION

The ROI is discretized using a raster grid with a resolution step of Δ=0.025 km (empirically chosen for operational EMSO level), defining a finite matrix of possible positions where the swarm of JXs can exist in. To stand for real EMSO deployment limits, the ROI is divided along the *y*-axis into two distinct operational zones:

- **Area of Interest (AOI)** is a foe territory and the part of ROI, resolute for deployment and maneuvering of one enemy GNSS RX:

$$\text{AOI} \cap \mathcal{A}, \forall x \ge 0 \quad (2)$$

- **Area of Responsibility (AOR)** is a friendly territory and the part of ROI, resolute for deployment and maneuvering of swarm of JXs:

$$\text{AOR} \cap \mathcal{A}, \forall x < 0 \quad (3)$$

The fixed GNSS RX is positioned at $\mathbf{x}_{\text{RX}} = [0, 5]$, while JXs can maneuver in the AOR. At time step *t*, the position of *i*-th (*i* =1,2, …, *N*=10) JX is $\mathbf{x}_{i,t} = [x_{i,t}, y_{i,t}]$. At the beginning of simulation (*t*=0), the JXs start the mission from separate positions and can move inside the AOR maneuvering boundary ($\forall x \le -5$). Such constraints show the real EMSO conditions and help to simulate the swarm jamming keeping the minimum distance from GNSS RX.

3.2. RADIO PROPAGATION MODEL

The research focuses particularly on GNSS L1 frequency band (1575.42 MHz), which power level ($P_{\text{GNSS,RX}}$) at the whole ROI randomly fluctuated in the range from -125 dBm to -130 dBm that is simply presented with Normal Gaussian distribution. The jamming signal arriving at the targeted GNSS RX position is modeled using a Free Space Path Loss (FSPL) propagation model:

$$L_i = 20\log_{10}(d_{i,t}) + 20\log_{10}(f) + 32.45 \quad (4)$$

where L_i is signal propagation loss from *i*-th JX to the targeted GNSS RX, $d_{i,t} = \|\mathbf{x}_{i,t} - \mathbf{x}_{\text{RX}}\|_2$ is the Euclidean 2D distance between the *i*-th JX and the targeted GNSS RX in km, *f* is GNSS RX frequency in MHz, and 32.45 is FSPL constant. To guarantee a successful EA, the cumulative Jamming-to-Signal Ratio (*JSR*) at the GNSS RX position must satisfy a predefined effectiveness threshold, which is set to γ_{JSR}=20.005 dB. The critical threshold value (γ_{JSR}) for effective EA can be observed through the cumulative power:

$$P^{\gamma}_{\text{J,RX}} = P_{\text{GNSS,RX}} + \gamma_{\text{JSR}} \quad (5)$$

$$p^{\gamma}_{\text{J,RX}} = 10^{\frac{P^{\gamma}_{\text{J,RX}}}{10}} \quad (6)$$

Next, assuming all JXs transmit with the same power, the required uniform power per JX is:

$$p_t^{\text{JX}} = \frac{p^{\gamma}_{\text{J,RX}}}{\sum_{i=1}^{N} g_i} \quad (7)$$

$$P_t^{\text{JX}} = 10\log_{10}(p_t^{\text{JX}}) \quad (8)$$

where P_t^{JX} is resulting jamming signal at the GNSS RX position and $g_i = 10^{(G_{JX}+G_{RX}-L_i/10)}$ is the linear path gain from *i*-th JX towards the GNSS RX, where G_{JX} is antenna gain of each JX, G_{RX} is antenna gain of GNSS RX and L_i is defined in equation (4). It is important to note that P_t^{JX} is power term used inside the reward function explained in chapter 3.1.

3.3. OPTIMIZATION VIA RL

Building upon the foundational concepts proved in our prior work [28], the problem of finding the best spatial distribution of the JXs is solved by engaging Proximal Policy Optimization (PPO) [29] due to its superior stability and clipped objective constraints in non-stationary multi-agent settings.

State Space (*s*). Comprises the current coordinates of all JXs in the AOR with the boundary restriction, alongside the current estimated *JSR* derived from direct feedback loops. The PPO agent was implemented using an actor–critic neural-network architecture. For each JX, the input state is a four-dimensional normalized vector composed of the current JX spatial coordinates, the required uniform JX power, and the estimated joint JSR at the GNSS RX:

$$s_{i,t} = \left[\frac{x_{i,t}}{D/2}, \frac{y_{i,t}}{D/2}, \frac{P_t^{\text{JX}}}{50}, \frac{JSR_t}{50}\right] \quad (9)$$

Action Space (*a*). Defines the permitted directional steps of the agent on the raster grid across the four primary basic directions (North, South, East, West) with a discrete step size of Δ_d, or a possibility to remain stationary:

$$a_{i,t} = [N(\uparrow), S(\downarrow), E(\rightarrow), W(\leftarrow), Stay(\cdot)] \quad (10)$$

Reward Function (*r*). The reward function was designed to guide the swarm of JXs toward spatial configurations that satisfy the required *JSR* at the GNSS RX while minimizing the required uniform JX power. Each JX receives a local reward composed of a shared power-efficiency term, a local distance-improvement term, a *JSR* tracking penalty, an invalid-action penalty, and a pairwise separation penalty. The power-efficiency term encourages configurations that reduce the uniform transmit power needed to achieve the target joint *JSR*. The distance-reduction term encourages each JX to move toward the RX within the allowed rasterized maneuvering region. The *JSR* tracking term penalizes deviations from the required *JSR*, while the invalid-movement action and JX-to-JX separation penalties enforce motion feasibility and

discourage unrealistic JX co-location. A compact form of the reward function is:

$$r_{i,t} = w_P(\Delta P_t) + w_d \Delta d_{i,t} - w_J e_{JSR,t} - w_B(I_{i,t}^{invalid}) - w_S(S_t) \quad (11)$$

where the power-efficiency reward is product of scaling factor (w_P=3.0) and power difference ($\Delta P_t = P_{t-1}^{JX} - P_t^{JX}$); the distance-reduction reward is product of scaling factor (w_d=0.75) and distance difference ($\Delta d_{i,t} = \frac{d_{i,t-1} - d_{i,t}}{\Delta}$); the *JSR* tracking penalty is product of scaling factor (w_J=0.05) and *JSR* mismatch ($e_{JSR,t} = \frac{|JSR_t - \gamma|}{\max(|\gamma|,1)}$); the invalid movement penalty is product of scaling factor (w_B=0.05) and invalid movement (when JXs try to leave the allowed maneuvering region); the JX-to-JX separation penalty is product of scaling factor (w_S=0.10) and JXs co-location or unwanted clustering. The scaling factors are empirically chosen during the training process.

The PPO network consists of a shared fully connected backbone with two hidden layers, each having 96 neurons and hyperbolic tangent (Tanh) activation functions. The actor head maps the hidden representation to logits over the five discrete raster actions (a). A categorical distribution is then formed from these logits to sample the JX movement action. In parallel, the critic head maps the same hidden representation to a scalar state-value estimate used for advantage estimation. The PPO network parameters are optimized using the Adam optimizer with a learning rate of $\alpha=3\times10^{-4}$. Training uses the clipped PPO objective (ε=0.2), generalized advantage estimation with ζ=0.97 and λ=0.95, a value-loss coefficient of 0.50, an entropy coefficient of 0.02, and gradient-norm clipping at 0.50. The PPO network training was implemented in PyTorch using a custom JX swarm environment rather than the OpenAI Gymnasium interface. Evaluation is performed with 1000 steps and greedy-polish mechanism. The proposed environment directly represents the multi-JX raster movement problem, RF link-budget evaluation, JSR computation, reward calculation, and movement constraints.

3.4. FL FRAMEWORK FOR JAMMING EFFECTIVENESS MONITORING

Centralized data aggregation between JXs in EMSO could be disrupted due to bandwidth constraints, communication masking, or the deployment of electronic counter-counter measurement (ECCM) assets. To mitigate this weakness, we propose an FL framework consisting of the distributed JXs as local FL clients. The purpose of the FL framework (FL-PPO) is to allow the JXs to improve their PPO policies without sending raw local trajectories, observations, rewards, or full training data to a central controller - aggregator. At the beginning of each federated round (r), the aggregator JX broadcasts the current global actor's weights ($\theta_i^{(r)} \leftarrow \theta^{(r)}$) to the other JXs. Each JX then performs local PPO training using its own local states, actions, rewards, and *JSR* feedback. After local training, i-th JX obtains updated local weights ($\theta_i^{(r,local)}$). Instead of sending the full local model or raw training data, each JX computes a compact weight increment. Moreover, only increment ($\Delta\theta_i^{(r)} = \theta_i^{(r,local)} - \theta^{(r)}$) is transmitted to the aggregator JX. Although the framework is federated, it still uses central aggregation. The difference from conventional architecture is that the central aggregator is one of the JX stations. The selected aggregator JX performs five main operations:

- **Global model initialization**. At round r=0, the aggregator JX initializes the global PPO actor model ($\theta^{(0)}$).
- **Weight-increment collection**. After local PPO training, each JX sends only its compact update ($\Delta\theta_a^{(r)}$) to the aggregator JX.
- **Compact update compression**. The weight increments are compressed before transmission with top-k FP16 compression.
- **Federated averaging**. The aggregator JX periodically executes a Federated Averaging routine to compute the updated global model parameters:

$$\Delta\theta_{golbal}^{(r)} = \frac{1}{N}\sum_{i=1}^{N} \Delta\theta_i^{(r)} \quad (12)$$

The aggregator JX reconstructs the sparse updates, computes their average, and the global model is then updated and broadcasted back to the JXs. This decentralized pipeline yields system robustness, enables collaborative observation of the ESM, and accelerates the convergence rate of the RL agents toward globally the best spatial coordination within highly fluid EMSO conditions.

4. RESULTS AND DISCUSSION

This section presents the simulation results, which were conducted to analyze the efficacy of jamming, and the prediction of the best locations for distributed JXs (two JXs, five JXs and ten JXs towards one GNSS RX). The results show that the proposed approach can be used for AI distributed jamming of GNSS RX in the context of EMSO. The FL-PPO framework can achieve stable convergence toward the best locations, with *JSR* values consistently supported within the specified threshold. Visual analysis of the agent's trajectories reveals that the model avoids EMS inefficient regions, focusing instead on positions that help the best *JSR*. Results of the optimal positions of two JXs with FL-PPO framework when GNSS RX is asymmetrically positioned are shown in Figure 1.

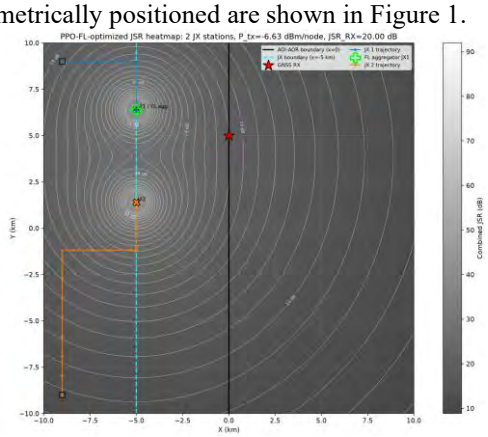


Figure 1. Optimal position of two JXs with FL-PPO when GNSS RX is asymmetrically positioned

The trajectories shown in Figure 1 demonstrate how the JXs focus on maneuvering toward the GNSS RX coordinates, optimizing their positions to maximize the *JSR* while minimizing transmit power. Results of the optimal positions of five JXs with FL-PPO framework

when GNSS RX is asymmetrically positioned are shown in Figure 2.

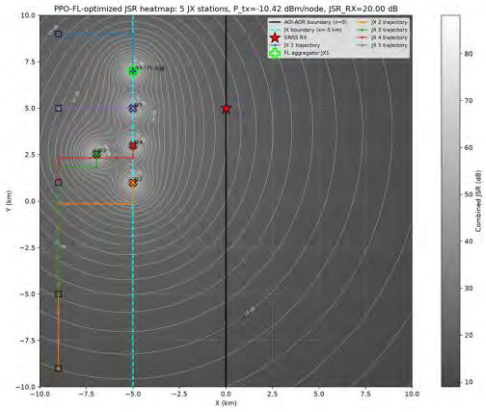
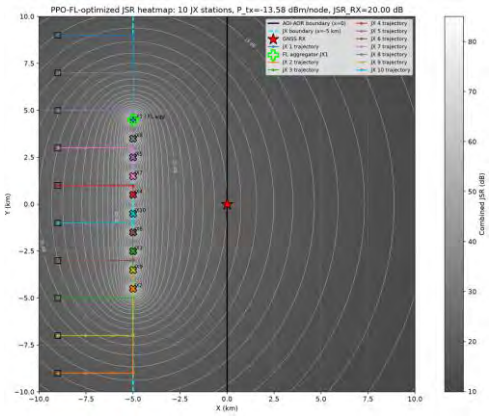
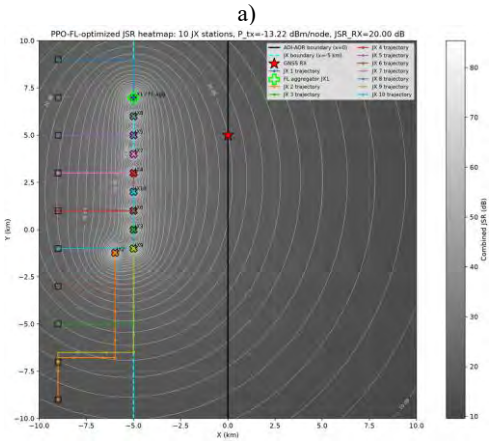


Figure 2. Optimal position of five JXs with FL-PPO when GNSS RX is asymmetrically positioned

It can be noticed that the higher density of JX trajectories occurs near the GNSS RX which confirms the FL-PPO framework adaptation when GNSS RX is asymmetrically positioned. Figures 1 and 2 show that the asymmetric setup shifts optimal JX locations toward the displaced GNNS RX, proving the proposed approach dynamically adapts to varying GNSS RX geometries rather than converging to fixed configurations. The improvement in the proposed approach can easily be seen with scenario when ten JXs are engaged, that is presented in Figure 3.



a)



b)

Figure 3. Optimal spatial distribution of ten JXs using FL-PPO: a) GNSS RX at the ROI center, and b) GNSS RX asymmetrically positioned at (0,5).

Figure 3 illustrates the optimized spatial deployment of ten JXs under two GNSS RX placement scenarios using the FL-PPO framework. It can be noticed that the spatial distribution contours confirm that individual JXs autonomously place themselves to avoid overlapping redundancies, ensuring a uniform *JSR* coverage over the GNSS RX position. During evaluations of edge cases (JXs were initialized at the border of the ROI), the FL-PPO framework successfully guided isolated JXs toward the globally optimal positions. The results demonstrate that the FL-PPO framework balance effectively from small configurations to larger JX's swarms, successfully satisfying joint *JSR* while minimizing transmit power through cooperative spatial positioning. These properties of the FL-PPO framework support findings and emphasize that RL models can outperform traditional heuristic and deterministic search algorithms in situations with expansive search domains or dynamically changing conditions [12], [22]. This can be supported by quantitative comparison shown in Table 1.

Table 1. Quantitative comparison.

Method	N JX	RX	P_t^{JX}/JX, [dBm]	Mean $d_{i,t}$ [km]	"greedy-polish" steps
PPO	2	(0,0)	-6.67	5.5901	18
		(0,5)	-6.39	6.0443	78
FL-PPO		(0,0)	-6.67	5.5917	63
		(0,5)	-6.63	5.6801	104
PPO	5	(0,0)	-10.59	5.7153	97
		(0,5)	-10.31	6.0822	221
FL-PPO		(0,0)	-10.59	5.7153	200
		(0,5)	-10.42	5.9085	206
PPO	10	(0,0)	-13.58	5.7332	146
		(0,5)	-13.17	6.2686	212
FL-PPO		(0,0)	-13.58	5.7330	180
		(0,5)	-13.22	6.1712	222

Table 1 demonstrates that FL-PPO framework achieves same or better spatial optimization and jamming effectiveness to centralized PPO, especially in the scenario when GNSS RX is asymmetrically positioned. Despite its advantages, the FL-PPO framework introduces a minor heterogeneity penalty. Specifically, in the ten JXs swarm scenario, the FL-PPO framework required 23% more total evaluation steps to achieve a stable spatial convergence when GNSS RX is symmetrically positioned, and 5% when GNSS RX is symmetrically positioned. Such delay in global model synchronization is an expected trade-off, attributed to the periodic nature of the federated weight aggregation phase compared to the continuous, immediate centralized updates.

5. CONCLUSION

This paper introduced a FL-PPO framework for autonomous distributed GNSS jamming using PPO-based local control and a peer-selected aggregation JX. By exchanging only compact model-weight increments rather than raw trajectory data, the architecture eliminates central dependence while maintaining communication efficiency. Simulation results confirm that both PPO and FL-PPO frameworks successfully adapt jammer deployment to target geometries—converging symmetrically for centered GNSS receivers and shifting dynamically toward asymmetric placements to minimize transmit power while

satisfying target JSR constraints. While the FL-PPO framework introduces a slight delay in convergence compared to a centralized PPO, it provides critical advantages in tactical EMSO: immunity to single-point communication failures and the ability to dynamically adapt to asymmetric target geometries. Furthermore, scalability tests demonstrate that the FL-PPO framework effectively coordinates larger swarms. The primary contribution lies in merging distributed spatial optimization with federated weight-sharing for communication-constrained environments. Future work will incorporate terrain-induced multipath propagation, dynamic jamming power and receiver mobility over predefined trajectory, network latencies, and hardware-in-the-loop validation.

Acknowledgements: This work was supported by the Ministry of Science, Technological development and Innovations of the Republic of Serbia, Contract No. 451-03-34/2026-03/200325.

References

[1] Joint Chiefs of Staff, *CJCS: Joint Publication 3-85: Joint Electromagnetic Spectrum Operations (JEMSO)*. 2022. Accessed: Jun. 28, 2026. [Online]. Available: https://www.waru.edu/sites/default/files/Migrated/CopDocuments/Joint_Pub_3-85.pdf

[2] V. Renaudin, M. I. Sayyaf, F. Le Bourhis, and M. Ortiz, "GNSS Positioning Under Threat: The Rising Risk to Existing Systems and The Role of Alternative Indoor and Seamless Navigation Technologies," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 3, pp. 246–259, Nov. 2025, doi: 10.1109/JISPIN.2025.3629705.

[3] O. Sushych, O. Pogurelskiy, V. Konin, O. Kutsenko, I. Prykhodko, and T. Maliutenko, "Performance Analysis of Intentional Interference on Multi-GNSS Receivers," *Lecture Notes in Networks and Systems*, vol. 992 LNNS, pp. 114–124, 2024, doi: 10.1007/978-3-031-60196-5_10/SAVE-RESEARCH.

[4] A. Blatnik and B. Batagelj, "Evaluating GNSS Receiver Resilience: A Study on Simulation Environment Repeatability," *Electronics 2025, Vol. 14, Page 1797*, vol. 14, no. 9, p. 1797, Apr. 2025, doi: 10.3390/ELECTRONICS14091797.

[5] B. Gattis, J. Cydejko, and D. Akos, "Baltic sea GNSS jamming and spoofing emitter detection and localization in real-time using a time difference of arrival (TDOA) system," *GPS Solutions*, vol. 30, no. 2, p. 96, Apr. 2026, doi: 10.1007/s10291-026-02061-5.

[6] Z. Meng and F. Shen, "Robust Space-Time Adaptive Processing Method for GNSS Receivers in Coherent Signal Environments," *Remote Sens. (Basel).*, vol. 15, no. 17, p. 4212, Aug. 2023, doi: 10.3390/rs15174212.

[7] J. W. Harris and D. M. Bevly, "Beam Steering Adaptive Noise Cancellation with a Controlled Reception Pattern Antenna for GNSS Anti-Jamming," in *2025 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, Apr. 2025, pp. 337–344. doi: 10.1109/PLANS61210.2025.11028269.

[8] A. M. D. Barbu, "The Cognitive Electronic Warfare in the Age of Artificial Intelligence," *STRATEGIES XXI - National Defence College*, vol. 2, no. 1, pp. 15–24, Nov. 2024, doi: 10.53477/2784-2487-24-02.

[9] T. Wang, Y. Li, Y. Wu, and T. Q. S. Quek, "Secrecy Driven Federated Learning via Cooperative Jamming: An Approach of Latency Minimization," *IEEE Trans. Emerg. Top. Comput.*, vol. 10, no. 4, pp. 1687–1703, Oct. 2022, doi: 10.1109/TETC.2022.3159282.

[10] A. Alhosban, "Electronic Warfare in NAVWAR: Impact of Electronic Attacks on GNSS / GBAS Approach Service Types C and D Landing systems and their proposed Electronic Protection Measures (EPM)," *Hadmérnök*, vol. 14, no. 2, pp. 238–255, Jun. 2019, doi: 10.32567/hm.2019.2.20.

[11] A. Meftah, T. N. Do, G. Kaddoum, C. Talhi, and S. Singh, "Federated Learning-Enabled Jamming Detection and Waveform Classification for Distributed Tactical Wireless Networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 4, pp. 5053–5072, Dec. 2023, doi: 10.1109/TNSM.2023.3271578.

[12] Z. A. El Houda, H. Moudoud, and B. Brik, "Federated Deep Reinforcement Learning for Efficient Jamming Attack Mitigation in O-RAN," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 9334–9343, 2024, doi: 10.1109/TVT.2024.3359998.

[13] Tony Anthony, "NAVSTAR GPS Space Segment/Navigation User Segment Interfaces," 2023.

[14] K. Špoljar, M. Zrinjski, A. Tupek, and K. Stipetić, "Modernization of GNSS, RNSS, and SBAS," *Technologies (Basel).*, vol. 13, no. 11, p. 494, Oct. 2025, doi: 10.3390/technologies13110494.

[15] S. Jin, X. Meng, G. Dardanelli, and Y. Zhu, "Multi-Global Navigation Satellite System for Earth Observation: Recent Developments and New Progress," *Remote Sens. (Basel).*, vol. 16, no. 24, p. 4800, Dec. 2024, doi: 10.3390/rs16244800.

[16] A. Mohanty and G. Gao, "A survey of machine learning techniques for improving Global Navigation Satellite Systems," *EURASIP J. Adv. Signal Process.*, vol. 2024, no. 1, p. 73, Jun. 2024, doi: 10.1186/s13634-024-01167-7.

[17] A. Morrison *et al.*, "Jammertest 2022: Jamming and Spoofing Lessons Learned," in *ENC 2023*, Basel Switzerland: MDPI, Oct. 2023, p. 22. doi: 10.3390/ENC2023-15445.

[18] L. Marata *et al.*, "Trends and Challenges in Next-Generation GNSS Interference Management," Oct. 2025, doi: 10.48550/arXiv.2510.27576.

[19] K. Radoš, M. Brkić, and D. Begušić, "Recent Advances on Jamming and Spoofing Detection in GNSS," *Sensors*, vol. 24, no. 13, p. 4210, Jun. 2024, doi: 10.3390/s24134210.

[20] A. Ghanbarzadeh, M. Soleimani, and H. Soleimani, "GNSS/GPS Spoofing and Jamming

Identification Using Machine Learning and Deep Learning," Jan. 2025. doi: 10.48550/arXiv.2501.02352.

[21] Y. K. Esfandabadi, A. Tabatabaei, and R. Hein, "AI-Driven Detection, Characterization and Localization of GNSS Interference: A Comprehensive Approach Using Portable Sensors," in *European Navigation Conference 2025*, Basel Switzerland: MDPI, Mar. 2026, p. 43. doi: 10.3390/engproc2026126043.

[22] S. You, M. Diao, and L. Gao, "Deep Reinforcement Learning for Target Searching in Cognitive Electronic Warfare," *IEEE Access*, vol. 7, pp. 37432–37447, 2019, doi: 10.1109/ACCESS.2019.2905649.

[23] X. Xiong *et al.*, "Intelligent jamming decision-making system based on reinforcement learning," *Computers and Electrical Engineering*, vol. 123, p. 110288, Apr. 2025, doi: 10.1016/j.compeleceng.2025.110288.

[24] W. Zhang, M. Zhang, Z. Zhao, F. Sun, and F. Liu, "Frequency jamming resource allocation method based on temporal reinforcement learning," *Signal Processing*, vol. 239, p. 110330, Feb. 2026, doi: 10.1016/j.sigpro.2025.110330.

[25] S. Barkatsa, M. Diamanti, P. Charatsaris, S. Voikos, E. E. Tsiropoulou, and S. Papavassiliou, "Coordinated Jamming and Poisoning Attack Detection and Mitigation in Wireless Federated Learning Networks," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 3745–3759, 2025, doi: 10.1109/OJCOMS.2025.3558672.

[26] S. Liu, W. Liu, A. Hussain, and P. Papadimitratos, "G2FL: Robust Federated Learning for GNSS Spoofing Detection," in *IEEE ICC 2026 Fifth Workshop on Machine Learning and Deep Learning for Wireless Security Proceedings and IEEE Xplore*,

[27] M. Jiang, Z. Ye, Y. Xiao, and X. Gou, "Federated Transfer Learning Aided Interference Classification in GNSS Signals," Jun. 2024, doi: 10.48550/arXiv.2406.16102.

[28] O. Pejčić, N. Šepa, and B. Sazdić-Jotić, "Predikcija efikasnosti radio-ometanja primenom algoritma veštačke inteligencije," in *Zbornik radova 69. konferencije ETRAN*, Društvo za ETRAN i Akademska misao, Beograd, Srbija, 2025, pp. 251–256. doi: 10.69994/69E25049.

[29] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Aug. 2017, doi: 10.48550/arXiv.1707.06347.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A COMPARATIVE EVALUATION OF SPECTROGRAM- AND BISPECTRUM- BASED TRANSFER LEARNING FOR DRONE AUDIO CLASSIFICATION

SARA MILANOVIĆ
Military Technical Institute, Belgrade, sara.milanovic@mod.gov.rs

MIODRAG VRAČAR
Military Technical Institute, Belgrade, mvracar566@gmail.com

IVAN POKRAJAC
Military Technical Institute, Belgrade, ivan.pokrajac@vs.rs

Abstract: *Acoustic signal analysis serves as an essential approach for detecting drones operated via optical fibers, which evade conventional radio-frequency detection methods. Nonetheless, achieving reliable acoustic-based detection poses notable challenges due to the pronounced sensitivity of classification models to environmental noise and low signal-to-noise ratios. While deep transfer learning has emerged as a powerful tool for audio classification, the choice of optimal time-frequency representation remains a critical factor in model performance. This paper presents a comparative evaluation of spectrogram- and bispectrum-based transfer learning for drone audio classification. Utilizing a comprehensive dataset of drone and non-drone acoustic signals, raw audio data is transformed into standard 2D spectrograms and higher-order bispectrum representations, which are then classified using three pre-trained models: ResNet-18, ResNet-50 and ViT Base 16. The results of this analysis suggest that higher-order spectral analysis provides a more robust and dependable feature representation in case of smaller model architectures, while larger networks struggle to generalize on the nonlinear features of the bispectrum and favor spectrogram representations.*

Keywords: *audio signal classification, drone classification, transfer learning, spectrogram, bispectrum.*

1. INTRODUCTION

The detection and classification of Unmanned Aerial Vehicles (UAVs), commonly known as drones, has become a matter of critical importance, namely for airspace surveillance and perimeter security. Existing counter-UAV systems employ several detection methods, including radar, radio-frequency (RF) monitoring, electro-optical and infrared imaging, and acoustic sensing. Each of these methods has its limitations – passive RF systems depend on the presence of detectable electromagnetic emissions and may be affected by interference from other wireless devices, whereas optical systems generally require an unobstructed view of the target and adequate illumination or atmospheric visibility [1]. These limitations motivate the investigation of complementary detection techniques.

The emergence of fibre-optic-controlled UAVs further illustrates the limitations of systems that depend on detecting or disrupting wireless command links. In such platforms, command and video signals are transmitted through a physical fibre-optic cable rather than a conventional radio link. Consequently, RF detectors designed to identify command-and-control transmissions may be ineffective [2]. Acoustic sensing is therefore of interest as a passive and comparatively low-cost method that does not depend on the UAV's communication protocol. It instead exploits the characteristic sounds generated by rotating propellers, electric motors, and associated mechanical components [3]. Its practical performance, however, is affected by propagation distance, microphone characteristics, wind, reverberation, and competing environmental sounds [4].

Machine Learning (ML)-based acoustic UAV detection commonly begins by converting a time-domain recording into a representation that exposes its spectral and temporal structure. Frequently used representations include short-time Fourier transform spectrograms, Mel-frequency spectrograms, filterbank energies, and Mel-frequency cepstral coefficients (MFCCs) [1], [3], [5]. A magnitude or power spectrogram describes the local distribution of signal energy across time and frequency, but it omits the phase of the complex STFT coefficients. In addition, additive noise contributes energy to the same time-frequency representation and may obscure the tonal and harmonic components associated with rotor and motor operation.

Higher-order spectral analysis offers a complementary representation of these signals. In particular, the bispectrum describes third-order statistical dependencies and can reveal quadratic phase coupling among frequency components [6], [7]. Such coupling may provide information about harmonic and nonlinear interactions that is not represented explicitly in an ordinary power spectrum. Moreover, for additive Gaussian noise that is statistically independent of the signal, third-order cumulants vanish in

Corresponding author: S. Milanović;
E-mail: sara.milanovic@mod.gov.rs
DOI: 10.5937/OTEH26059M

expectation. Bispectral estimation can therefore reduce the expected contribution of Gaussian noise [6], [7]. Most frequently used methods for estimating this bispectral structure include parametric approaches based on autoregressive moving-average (ARMA) models [8].

The selection of a suitable classifier is also influenced by the quantity and diversity of the available training data. Convolutional Neural Networks (CNNs), including residual architectures such as ResNet, incorporate local spatial structure through convolution and have been widely adopted for time-frequency image inputs [9]. Transformer architectures, including the Vision Transformer and the Audio Spectrogram Transformer, use self-attention to model relationships between patches or time-frequency regions over longer ranges [5], [10]. However, the original success of vision transformers depended substantially on large-scale pretraining, and training such models from scratch can be difficult when only a limited dataset is available. Data-efficient training, transfer learning, augmentation, and regularization can mitigate this limitation [10], [11]. Thus, model performance in small datasets depends not only on architecture capacity, but also on initialization and training strategy.

2. LITERATURE

Acoustic UAV detection has predominantly been formulated as an environmental sound-classification problem using short-time spectral representations. Jeon et al. [12] compared Gaussian mixture models, CNNs, and Recurrent Neural Networks (RNNs) using MFCC and Mel-spectrogram features, including evaluation with recordings made in an urban street environment. Bernardini et al. [13] adopted a more traditional pipeline, combining temporal and spectral descriptors—such as short-time energy, spectral centroid, spectral roll-off, and MFCCs—with a support vector machine (SVM) model. Al-Emadi et al. [14] evaluated CNN, RNN, and Convolutional Recurrent Neural Network (CRNN) architectures using spectrogram representations and investigated synthetic data generation as a means of enlarging the available UAV-audio dataset. Other studies have explored model and feature fusion: Casabianca and Zhang [15] combined the outputs of multiple deep networks operating on Mel-spectrograms, whereas Dong et al. [16] fused decisions obtained from Log-Mel-spectrogram and MFCC-based classifiers. Collectively, these studies demonstrate the feasibility of learning drone acoustic signatures from time-frequency representations.

Nevertheless, the robustness of acoustic UAV detectors remains sensitive to propagation distance, environmental noise, microphone characteristics, UAV type, and differences between training and deployment conditions. Tejera-Berengué et al. [4] explicitly examined the effects of distance and recording environment on several conventional and deep-learning classifiers. More recent work has similarly emphasized cross-dataset generalization and the value of pretraining and noise augmentation [17]. Thus, it is safe to say that many published results are obtained from relatively limited, augmented, or closely matched datasets, and that reliable performance in previously unseen acoustic environments remains an open challenge.

As mentioned, most existing acoustic UAV classifiers rely on second-order signal representations (power spectra, STFT spectrograms, Mel-spectrograms), while bispectral representations have been investigated primarily for radar and RF signal classification rather than acoustic detection. For example, Liu et al. [18] combined bispectral slices with a neural classifier to recognize low signal-to-noise (SNR) UAV radar-signal modulations. In the available literature analyzed for this study, no peer-reviewed work was found that combines UAV acoustic bispectra with pretrained vision architectures. Parametric bispectrum estimation is itself well established. Raghuveer and Nikias developed an autoregressive approach in which model parameters are estimated from third-order recursion equations, providing a parametric alternative to conventional Fourier-based bispectrum estimators [5].

3. METHODOLOGY

In this section, we describe the framework utilized to classify drone audio signals and evaluate the performance of the model. Section 3.1 describes the dataset used in this study, along with the necessary preprocessing steps. Following the data description, Section 3.2 introduces the feature extraction pipeline. We give a brief explanation of spectrograms and bispectral representations of audio signals, used to capture temporal, spectral and higher-order statistical variations. In Section 3.3, the process of model training and evaluation is outlined, focusing on the method of transfer learning and the implemented cross-validation evaluation strategy.

3.1 Dataset

The dataset used for this research was taken from the HuggingFace hub [19]. It consists of ~180k .wav files with sampling rate of 16kHz, and each file belongs to one of the two categories: DRONE or NO_DRONE. Because of the substantial volume of the dataset and the resulting computational costs, we extracted an approximately balanced subset for the purpose of this study.

Since the original files are of varying lengths, they were segmented into 1-second snippets to satisfy the fixed-input requirements of neural network training process. This can result in the appearance of snippets containing no valuable information, which is why the resulting dataset was filtered using a percentile-based root mean square (RMS) energy threshold. With this, we acquired a subset consisting of ~3500 audio signals.

As it can be concluded from the class names, the DRONE class includes audio signals which contain various drone sounds in different operating environments. The samples were aggregated from multiple data sources, primarily capturing the acoustic profiles of various multirotor UAVs, with commercial quadcopters being the most dominant. On the other hand, the audio files in the NO_DRONE class represent a variety of sounds, including traffic, animal noises, chatter of people, etc. and were specifically pulled from environmental audio datasets. In some of the files, multiple different sounds are combined. For example, audio files from DRONE class may include distant traffic,

wind and outdoor ambient noise in the background.

3.2 Feature extraction

One of the most important steps before starting the ML model training process is selecting the right features to appropriately represent the data we want to classify. When working with audio files, there is a possibility of extracting features from the time domain (mean, standard deviation, zero-crossing rate, etc.), but this has ultimately proven to be inferior to the frequency-domain approach.

Figures 1 and 2 show examples of spectra of signals from our two different classes. As we can see in the DRONE signal spectrum, there is a series of regularly spaced, distinct peaks which correspond to the harmonic comb structure. This pattern can be attributed to the drone's electric motors and rotating propeller blades.

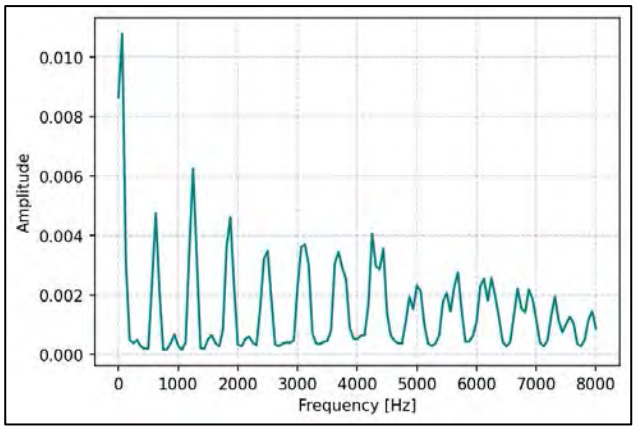


Figure 1. Spectrum of a DRONE signal

The difference between these two types of signals is clear when comparing figures 1 and 2. The NO_DRONE signal's energy is concentrated in the lower frequencies and undergoes a rapid decay past 1000Hz. The lack of mechanical periodicity here is obvious and the spectrum's appearance is highly characteristic of ambient background noise.

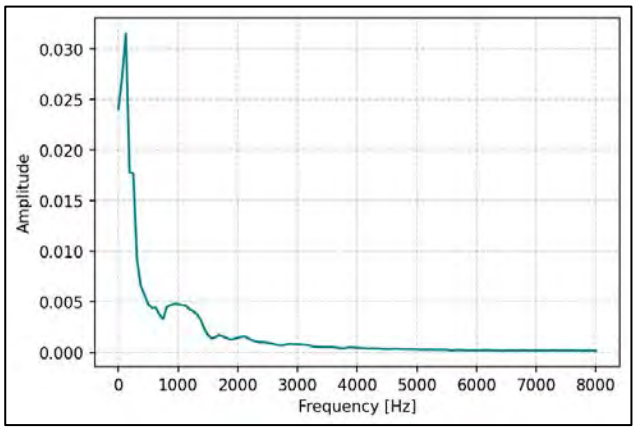


Figure 2. Spectrum of a NO_DRONE signal

Spectrogram is one of the most commonly used time-frequency representations of a signal. We use the Discrete Short-Time Fourier Transform (DSTFT) to calculate the frequency-domain representation of an audio file, given by the formula:

$$X[m,k] = \sum_{n=0}^{N-1} x[n+mR]w[n]e^{-j\frac{2\pi}{N}kn} \qquad (1)$$

In equation (1), $X[m,k]$ is the complex-valued STFT coefficient at time index m and frequency bin k, $x[n+mR]$ is the $(n+mR)$-th sample of the original audio sequence, $w[n]$ is the window function, R is the hop size, and N is the total number of samples. The spectrogram can then be calculated as:

$$S[m,k] = |X[m,k]|^2 \qquad (2)$$

While they are highly effective for extracting stationary characteristics, they cannot fully capture complex, non-linear dynamics often present in audio signals.

Figures 3 and 4 show examples of spectrograms of signals from the two different classes.

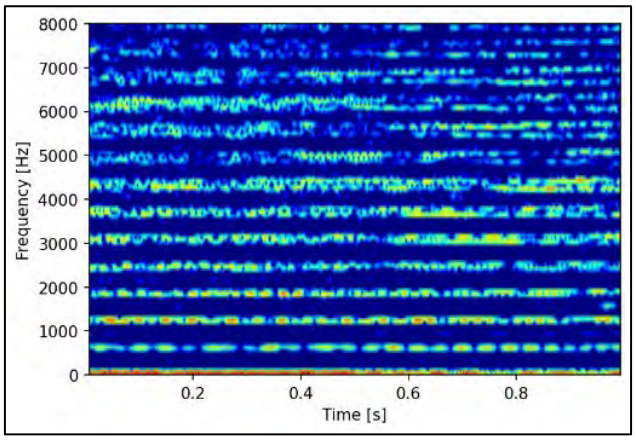


Figure 3. Spectrogram of a DRONE signal

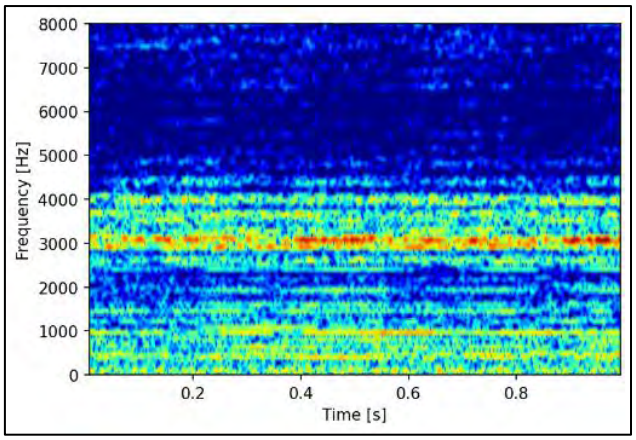


Figure 4. Spectrogram of a NO_DRONE signal

From these two representations, we can draw the same conclusions as we did when looking at figures 1 and 2. In the case of DRONE signal, we can see multiple "stripes" in the spectrogram which are the result of drone's motors and rotating blades. On the other hand, the NO_DRONE signal spectrogram has no such patterns and its values are noticeably lower for frequencies past ~4kHz.

Because of spectrogram's limitations, we calculated the bispectrum (third-order cumulant spectrum). The advantage of the bispectrum is in its ability to preserve phase information and detect non-linear interactions

between frequency components. Mathematically, the bispectrum is defined as the Fourier transform of its third-order cumulant sequence. Again, we use the DFT to practically calculate the bispectrum of a signal:

$$B[k_1, k_2] = E\left[X[k_1]X[k_2]X^*[k_1 + k_2]\right] \quad (3)$$

where $X[k1]$ and $X[k2]$ are the DFT coefficients at the frequency bins $k1$ and $k2$, respectively, X^* is the complex conjugate of the DFT coefficient at the sum $[k_1 + k_2]$ frequency bin, and E is the expectation operator. Because the resulting bispectrum is complex-valued, we calculate the absolute magnitude to extract a real-valued matrix suitable for neural network training.

Alternatively, the bispectrum can be calculated using Autoregressive Moving Average (ARMA) coefficients, and this is a parametric method that we used for this research. The method relies on modeling the audio signal as an ARMA process, estimating the coefficients and calculating the transfer function. Without going into the mathematics of this approach, the bispectrum is calculated as:

$$B[k_1, k_2] = \gamma_3 H(k_1)H(k_2)H^*(-k_1 - k_2) \quad (4)$$

In the previous equation, H and γ_3 are the transfer function and parameter obtained from the ARMA estimation. Determining the ARMA model order is tricky, but if done right, the resulting bispectrum will have higher frequency resolution and less smoothing issues (which are inherent to windowing functions) than is the case when calculating spectrograms.

For the research done in this paper, we used MatLAB to perform automatic ARMA model order selection [20] and the corresponding bispectrum matrices calculation, which we then loaded into Python prior to the model training. Spectrograms were calculated directly in Python. Figures 5 and 6 show examples of the resulting bispectra.

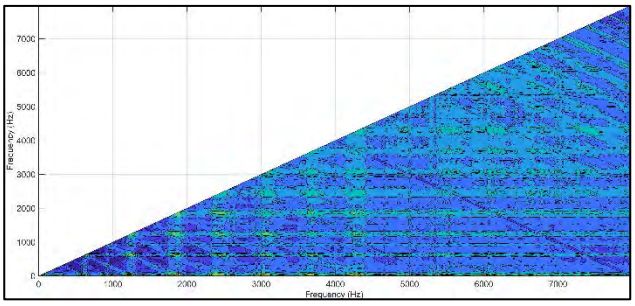


Figure 5. Bispectrum of a DRONE signal

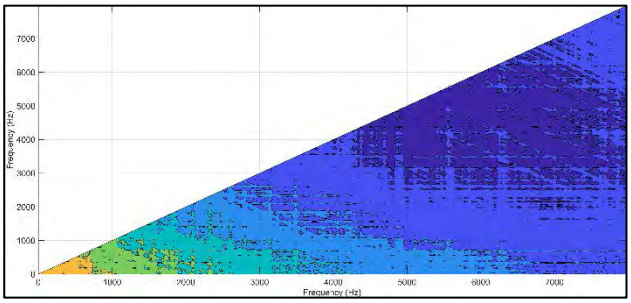


Figure 6. Bispectrum of a NO_DRONE signal

As can be seen in figures 5 and 6, because of bisepctrum's symmetrical nature and in order to lower the computational costs, we only kept the right-triangular matrix information for each bispectrum. Also, the bispectral magnitudes were converted to a logarithmic scale, which provided better results than linear scaling.

In these figures we can notice the same acoustic signatures we observed by analyzing the spectra. When looking at the bispectrum of a DRONE signal, we notice the intersecting vertical and horizontal lines which effectively form a grid pattern. This is the result of drone's deterministic, mechanical parts – rotating motors and blades.

The NO_DRONE signal's bispectrum shows strong phase coupling only in the low-frequency range. Signals from this class include various environmental sounds that are highly localized in both time and frequency, which explains why they are present as small, isolated phase coupling clusters in the bispectrum. As they lack the deterministic nature of drone signals, it makes perfect sense that there are no grid-like patterns visible, with only the dark blue color dominating the high-frequency range.

3.3 Model training and evaluation

As mentioned before, we opted for transfer learning with three pretrained models for our analysis: ResNet-18, ResNet-50 [9] and ViT Base 16 [10]. Instead of training a completely new neural network from scratch, transfer learning refers to using an already trained network and slightly adjusting it for a specific task. With this approach, we can be sure that the model has already learned common important features and we only need to fine-tune it, which also significantly reduces the amount of parameters it needs to learn and the resources it uses during that process.

Expectedly, this then poses a question: how many layers should we retrain (fine-tune) when using transfer learning? The models we selected vary significantly in the number of trainable parameters. ResNet-18 and ResNet-50 consist of roughly 12M and 26M parameters, respectively. On the other hand, ViT Base 16 is considerably larger, including around 85M trainable parameters. Considering the size of our dataset, we opted for the next approach: for ResNet-18 and ResNet-50, we made only deeper residual blocks trainable (layer3 and layer4 in PyTorch implementation). In case of ViT Base 16 model, the final six encoder blocks (blocks 6-11) were unfrozen for fine-tuning. The final fully connected layer is the same in all three cases — it includes a layer mapping the backbone features to a 128-dimensional space, followed by a ReLU activation function and a dropout layer ($p = 0.4$) to combat overfitting. For the optimizer we used AdamW, with learning rates ranging from 10^{-3} to 10^{-5}, depending on the depth of the layers.

One of the biggest concerns when it comes to training neural networks is their resulting generalization ability. It is not correct to focus solely on training accuracy and loss, because of possible overfitting. That is why we split the entire dataset into training and validation subsets and monitor the model's performance on both. To get an even better idea of our model's ability to generalize, we can use k-fold cross-validation. This method involves splitting the dataset into k subsets, using one subset for validation and the remaining $(k - 1)$ subsets for training. This process is

repeated k times and the resulting metrics are averaged. In this paper, we use 10-fold cross-validation (90% of the data is used for training and 10% for validation) to calculate our model's accuracy on the OOF (out-of-fold) samples (samples that were in the validation subset in each fold). Furthermore, the technique called Early Stopping is used during training in each fold. We observe validation loss with the patience parameter set to 3, meaning that if the loss did not achieve a new minimum after 3 epochs, the training process is terminated and the last best model weights are recorded and used for evaluation.

4. RESULTS

In this section, we evaluate the classification performances of our three selected models across two distinct feature extraction paradigms: the non-parametric STFT spectrogram and the parametric ARMA bispectrum. All models were trained and evaluated using the partial fine-tuning technique and 10-fold cross-validation, as described in the previous section. Because we used cross-validation for model training and evaluation, classification performance is quantified with mean accuracy (along with standard deviation), calculated on the OOF (out-of-fold) samples. Table 1 presents this metric for all mentioned experimental configurations.

Table 1. Classification results

	ResNet-18	ResNet-50	ViT Base16
Spectrogram	96.66 (±1.20)	96.46 (±0.89)	95.50 (±0.92)
Bispectrum	**96.92 (±0.78)**	96.05 (±1.19)	93.20 (±1.93)

We can draw several interesting conclusions looking at the results presented in Table 1. Firstly, ResNet-18 achieved the best classification performance, with bispectrum representations as inputs. The lowest standard deviation also makes this model the most consistent. This is particularly interesting because ResNet-18 has the smallest architecture here.

On the other hand, while it yields worse results than ResNet-18, ResNet-50 outperforms ViT Base 16 for both feature extraction methods. Especially intriguing is the fact that both ResNet-50 and ViT Base 16 achieve higher accuracy with spectrogram inputs. It could be argued that one of the reasons for these outcomes is the size of the dataset. An insufficient number of samples can cause models with large number of parameters (like ResNet-50 and especially ViT) to start learning specific details in training data that are not descriptive and/or represent noise. This ultimately leads to worse generalization ability and can only be remedied by increasing the dataset size. The explanation why larger models favor spectrograms in this experiment is not easy to explain. Spectrograms preserve the temporal information of the signal, while the ARMA bispectrum completely eliminates the time component in favor of focusing on frequency interactions. Because of their complex architectures and massive parameter numbers, large models may end up struggling trying to generalize on the bispectrum's nonlinear features.

We can take a look at the confusion matrix and ROC curve for the best-performing model, ResNet-18. Figure 7 shows the confusion matrix for said model.

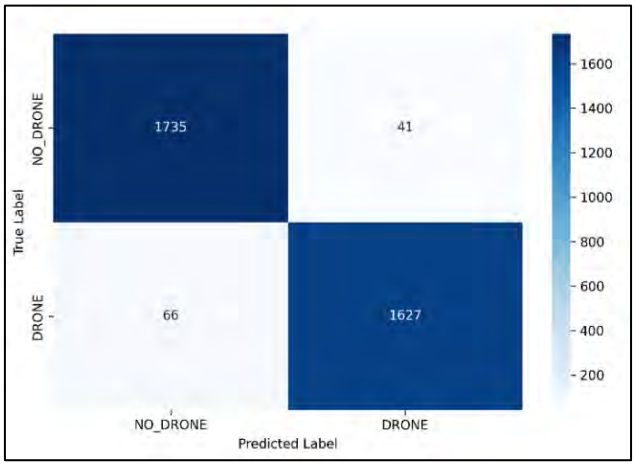


Figure 7. Classification matrix for ResNet-18

The matrix was obtained by aggregating all the OOF predictions, which is why the number of samples in the matrix is equal to the dataset size. As can be seen from the figure, classification performance is roughly symmetrical; the model achieves high precision and recall, indicating that there is no bias toward one specific class. The number of False Negative cases is higher than the number of False Positives, which is unfavorable for this specific task of drone classification.

Figure 8 shows the ROC curve for the best-performing model. The model achieved an Area Under the Curve (AUC) of 0.994. Because this metric was calculated based on out-of-fold (OOF) predictions across all ten folds, it provides a robust estimate of the model's classification performance and generalization ability.

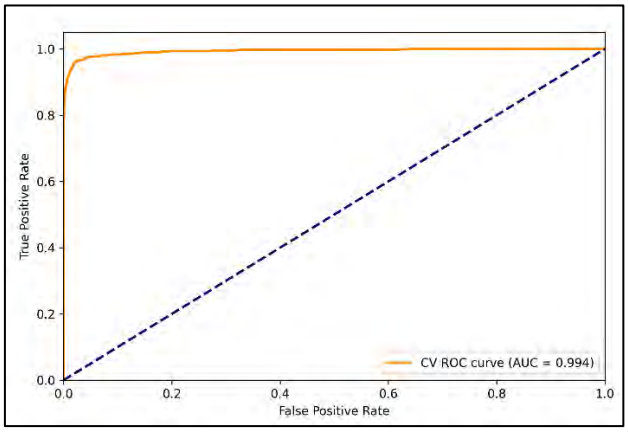


Figure 8. ROC curve for ResNet-18

5. CONCLUSION

In this paper, we performed a comparative evaluation of various transfer learning paradigms, focusing on audio drone classification. We utilized three different model architectures – ResNet-18, ResNet-50 and ViT Base 16. The networks were cross-examined across two distinct feature spaces: non-parametric STFT spectrograms and parametric ARMA bispectra.

The experiments revealed some unexpected results. The smallest model, ResNet-18, consistently outperformed

larger models, in both feature spaces. This best-performing model yields a ROC curve with the AUC of 0.994, with symmetrical classification performance. Non-parametric STFT spectrograms proved more valuable for larger networks, which is especially evident in the case of ViT Base 16. This model achieves a significantly higher accuracy when accepting spectrograms as inputs than is the case with bispectrum data.

It should be mentioned that the primary limitation of current research is the size of the dataset used. For large model architectures (notably ViT Base 16 in this case), data count is of critical importance. Consequently, the logical next step in this research would be the expansion of the dataset. With this, new experiments can be performed, which would help in determining whether large deep learning models can learn to better generalize on ARMA-calculated bispectrum inputs.

Based on the currently obtained results, another research direction involves investigating the usage of smaller model architectures (because of their superior performance) and performing testing with audio signals of drones operating in real-world environments.

References

[1] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," *Sensors*, vol. 24, no. 1, art. 125, 2024.

[2] A. Braszko, *Fiber Optic Drones: Posing a Significant C-UAS Challenge*, Center for Army Lessons Learned, No. 25-1046, 2025.

[3] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, "Audio Based Drone Detection and Identification Using Deep Learning," in *2019 15th International Wireless Communications and Mobile Computing Conference*, pp. 459–464, 2019.

[4] D. Tejera-Berengué, F. Zhu-Zhou, M. Utrilla-Manso, R. Gil-Pita, and M. Rosa-Zurera, "Analysis of Distance and Environmental Impact on UAV Acoustic Detection," *Electronics*, vol. 13, no. 3, art. 643, 2024.

[5] Y. Gong, Y.-A. Chung, and J. Glass, "AST: Audio Spectrogram Transformer," in *Proceedings of Interspeech 2021*, pp. 571–575, 2021, doi: 10.21437/Interspeech.2021-698.

[6] C. L. Nikias and M. R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework," *Proceedings of the IEEE*, vol. 75, no. 7, pp. 869–891, 1987.

[7] J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, 1991.

[8] M. R. Raghuveer and C. L. Nikias, "Bispectrum Estimation: A Parametric Approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 1213–1230, 1985.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[10] A. Dosovitskiy et al., "An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations*, 2021.

[11] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training Data-Efficient Image Transformers and Distillation through Attention," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, pp. 10347–10357, 2021.

[12] S. Jeon et al., "Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks," in *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, 2017.

[13] A. Bernardini, F. Mangiatordi, E. Pallotti, and L. Capodiferro, "Drone Detection by Acoustic Signature Identification," *Electronic Imaging*, pp. 60–64, 2017.

[14] S. Al-Emadi, A. Al-Ali, and A. Al-Ali, "Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks," *Sensors*, vol. 21, no. 15, art. 4953, 2021.

[15] P. Casabianca and Y. Zhang, "Acoustic-Based UAV Detection Using Late Fusion of Deep Neural Networks," *Drones*, vol. 5, no. 3, art. 54, 2021.

[16] Q. Dong, Y. Liu, and X. Liu, "Drone Sound Detection System Based on Feature Result-Level Fusion Using Deep Learning," *Multimedia Tools and Applications*, vol. 82, pp. 149–171, 2023.

[17] J. Reuter, H. Ohlenbusch, and S. Rollwage, "Improving Acoustic Drone Detection Generalization through Pretraining and Data Augmentation," arXiv:2605.31329, 2026.

[18] Liu et al., "Modulation Recognition of Low-SNR UAV Radar Signals Based on Bispectral Slices and GA-BP Neural Network," *Drones*, vol. 7, no. 7, art. 472, 2023.

[19] GeronimoBasso. (n.d.).: *Drone audio detection samples (DADS) [Dataset]*. HuggingFace. https://huggingface.co/datasets/geronimobasso/drone-audio-detection-samples.

[20] P. M. T. Broersen. ARMAsel for Irregular or Missing Data (https://www.mathworks.com/matlabcentral/fileexchange/18429-armasel-for-irregular-or-missing-data), MATLAB Central File Exchange. Retrieved August 5, 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

IMPACT OF JPEG COMPRESSION ON YOLO-BASED ROAD SCENE OBJECT DETECTION IN VISIBLE AND THERMAL IMAGES

BOBAN PAVLOVIĆ
University of Defence in Belgrade, Military Academy, boban.pavlovic@va.mod.gov.rs

BOBAN BONDŽULIĆ
University of Defence in Belgrade, Military Academy, boban.bondzulic@va.mod.gov.rs

DIMITRIJE BUJAKOVIĆ
University of Defence in Belgrade, Military Academy, dimitrije.bujakovic@va.mod.gov.rs

MOMIR STANKOVIĆ
University of Defence in Belgrade, Military Academy, momir.stankovic@va.mod.gov.rs,

Abstract: *Object detection is a fundamental component of road-scene perception systems used in autonomous driving, where visual data are often compressed to meet storage and bandwidth constraints. While image compression reduces transmission costs, it may introduce distortions that adversely affect detection accuracy. This study investigates the impact of JPEG compression on seven recent YOLO models (YOLOv5m, YOLOv7, YOLOv8m–YOLOv12m), evaluated on visible (VIS) and long wave infrared (LWIR) images from the M^{3}FD dataset. A balanced subset containing daytime and nighttime scenes is used, and ten compressed versions are generated with quality factor (QF) values ranging from 50 to 5. Results show that YOLOv12m achieves the highest global performance for VIS images, particularly during the day, while YOLOv10m excels under low-light conditions. For LWIR data, YOLOv8m performs best overall, with YOLOv12m showing advantages at night. Detection performance decreases gradually for QF values from 50 to 20, but drops sharply below QF=20, highlighting the importance of maintaining moderate image quality. Across intersection over union (IoU) thresholds, LWIR maintains more stable mAP than VIS, demonstrating its robustness under varying illumination. These findings provide quantitative insights into how JPEG compression and imaging modality affect modern YOLO detectors, informing reliable deployment under bandwidth and storage constrained environments.*

Keywords: *road scene object detection, multispectral image, JPEG compression, YOLO.*

1. INTRODUCTION

Autonomous driving systems depend on accurate object detection to perceive complex road scenes, including pedestrians, vehicles, and other traffic participants [1]. These systems rely on visual data for safe navigation and real-time decision making under diverse environmental and lighting conditions. The growing deployment of intelligent transportation systems has vastly increased visual data volumes, straining storage and transmission bandwidth. To mitigate this, image compression techniques such as JPEG are widely used, albeit at the cost of introduced distortions that may affect downstream tasks [2]. Deep learning-based object detection has become a cornerstone of computer vision, enabling machines to locate and classify objects of interest [3]. Because detection is directly tied to image content, compression-induced artifacts can alter edges, textures, and fine details critical for reliable recognition, particularly in adverse conditions like low illumination.

In recent years, YOLO (You Only Look Once) models have achieved notable success across computer vision tasks, including detection, segmentation, and pose estimation [4]. They offer an excellent speed-accuracy trade-off via single-pass inference. YOLO is widely adopted in surveillance, autonomous driving, robotics, medical imaging, and augmented reality [5]. The family has evolved through multiple generations, each addressing prior limitations through architectural innovations, parameter optimization, and improved training methodologies [6], yielding substantial gains in accuracy and efficiency while preserving real-time performance.

Numerous studies have assessed JPEG compression's impact on deep learning-based detection and classification. Dodge and Karam [7] examined effects of Gaussian blur, noise, contrast changes, and JPEG/JPEG2000 compression on classification models, finding CNNs particularly sensitive to blur and noise, with notable accuracy degradation. In [8], authors showed that compression artifacts degrade CNN-based object detectors' accuracy. Similarly, [2] analyzed JPEG compression on RetinaNet and Faster R-CNN, reporting pronounced detection drops at high compression levels.

Bayerl et al. [9] studied remote sensing image compression (JPEG, JPEG2000) on classification, showing that mismatched compression settings between

Corresponding author: B. Pavlović;
E-mail: boban.pavlovic@va.mod.gov.rs
DOI: 10.5937/OTEH26060P

training and inference cause substantial accuracy loss. In [10], authors evaluated HEVC compression on object tracking using YOLOv3 and SORT, finding that quantization significantly affects tracking performance while motion search range does not, and derived a regression model linking compression strength to accuracy. In [11], the robustness of several YOLO detectors to common distortions in remote sensing was evaluated, including Gaussian noise, JPEG/JPEG2000 compression, and blur. All distortions degraded accuracy, with YOLOv9 most robust and YOLOv7 and YOLOv10 most sensitive.

Despite these efforts, the robustness of YOLO detectors on infrared imagery, especially compared to visible (VIS) images under compression, remains insufficiently explored. Most existing studies focus on RGB data, leaving JPEG compression effects on long-wave infrared (LWIR) largely unexamined. Moreover, comparative VIS–LWIR analyses are limited, despite their distinct statistical and perceptual properties, which may lead to different responses to compression artifacts affecting textures and object boundaries. To fill this gap, this paper analyzes seven YOLO models – YOLOv5m [12], YOLOv7 [13], YOLOv8m [4], YOLOv9m [14], YOLOv10m [15], YOLOv11m [6], and YOLOv12m [16] – using a balanced test set from the M^3FD dataset, containing equal VIS and LWIR images split evenly between daytime and nighttime conditions. Both modalities are compressed with JPEG quality parameters from 50 down to 5 to systematically evaluate compression impact on detection performance.

The remainder of this paper is structured as follows. Section 2 describes the methodology, covering the dataset, YOLO models, and JPEG compression process. Section 3 presents experimental results with analysis and discussion. Section 4 summarizes key findings and outlines future research directions.

2. METHODOLOGY

2.1. Dataset description

In this study, VIS and LWIR images were used from the M^3FD [17] dataset, which is widely adopted for evaluating object detection performance in driving scenarios. A total of 440 VIS images and 440 LWIR images were selected to represent diverse driving environments. For each modality, 220 images correspond to daytime scenes and 220 images to nighttime conditions, ensuring a balanced evaluation across lighting scenarios. The selected images include typical road scenes containing pedestrians (referred to as person), cars, and buses, which are the three object categories considered in this study. The selected set covers both urban and suburban environments and includes variations in object scale, occlusion, and background complexity, providing a challenging and realistic benchmark for assessing detection performance under compression.

Table 1. Distribution of object instances in the selected subset across daytime, nighttime, and combined scenarios

Class	Daytime vision	Nighttime vision	All-day/night vision
Person	117	206	323
Car	219	220	439
Bus	147	10	157
All	**483**	**436**	**919**

Figure 1 shows images from both modalities - visible and infrared based on both daytime and nighttime conditions. The selected examples illustrate how a single image can contain representatives of multiple classes, as well as multiple samples within the same class. Additionally, it can be observed that objects may be close or far away, which would test the robustness of the detectors themselves. Objects may be visible in one modality but not visible in the other (see the presence of a person in nighttime conditions in LWIR, and their absence in VIS).

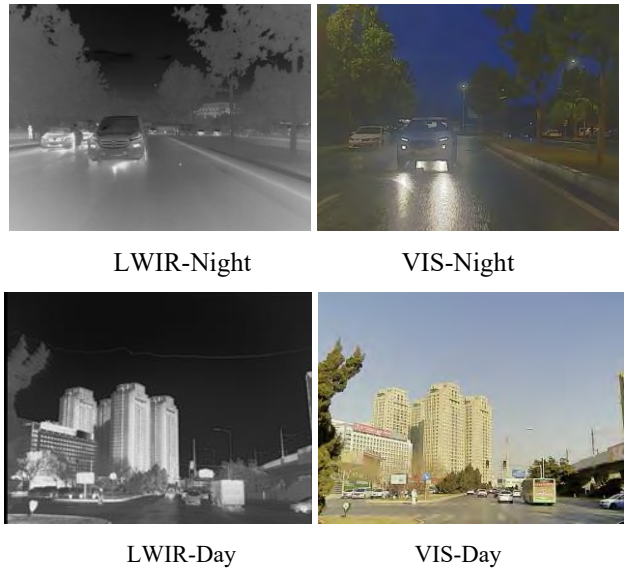


Figure 1. Example images from visible and infrared modalities based on daytime and nighttime conditions

2.2. YOLO models

Seven YOLO models are selected in this study (YOLOv5, YOLOv7–YOLOv12) to provide a comprehensive comparison across different design stages and performance trade-offs. YOLOv5m [12] introduced a stable anchor-based design using CSPDarknet53 with FPN and PANet for multi-scale features extraction. YOLOv7 [13] enhanced this design with E-ELAN and improved scaling strategies to better balance accuracy and speed. YOLOv8m [4] adopted an anchor-free framework, incorporating C2f modules and a decoupled head to improve gradient flow and detection precision. YOLOv9m [14] further improved training stability through programmable gradient information and employed the GELAN backbone for higher efficiency. YOLOv10m [15] focused on real-time performance by introducing a fully NMS-free pipeline with dual label assignment to reduce latency while improving accuracy. YOLOv11m [6] extended these developments with advanced attention mechanisms, such as C2PSA, C3K2 modules, and the SPPF block to better capture spatial

context while maintaining fast inference. The latest version, YOLOv12m [16], continues this evolution by refining architectural efficiency and training strategies to improve accuracy, robustness, and deployment flexibility.

2.3. JPEG compression

JPEG is a widely used lossy image compression method that reduces image size while maintaining acceptable visual quality [18]. It works by transforming the image into the frequency domain using the Discrete Cosine Transform (DCT), which separates visual information into low- and high-frequency components. Low frequencies capture most of the image structure, while high frequencies represent fine details and noise. The resulting DCT coefficients are then quantized to remove perceptually less important information, achieving significant data reduction. JPEG is valued for its strong balance between compression efficiency, visual fidelity, and low computational cost, enabling fast encoding and decoding. The compression level is controlled by a quality factor (QF) ranging from 1 to 100, where higher values correspond to lower compression and better image quality. In this study, JPEG compression was applied at 10 levels by varying the QF from 5 to 50 in steps of 5 using the MATLAB *imwrite* function.

3. RESULTS AND DISCUSSION

3.1 Experimental setup

The experimental setup was conducted on a Windows 11 operating system, running on an Intel® Core™ i7-12650H processor (2.30 GHz) with 16 GB of RAM, and an NVIDIA RTX 3050Ti GPU with 4 GB of memory. All algorithms were implemented using the PyTorch framework (version 2.0.1) and accelerated with the CUDA Toolkit (version 11.8).

In this study, all YOLO models are evaluated using pretrained weights from the MS-COCO dataset [19], which contains 80 object classes, including those selected from the M^3FD dataset. Each model is tested on the original image subsets as well as on ten JPEG-compressed versions generated by varying the QF. The evaluations are conducted separately for VIS and LWIR modalities and further divided into daytime and nighttime conditions to analyze the influence of illumination, providing a comprehensive comparison of detection performance across compression levels, imaging modalities, and lighting scenarios.

To assess the robustness of the YOLO models, widely adopted evaluation metrics were employed, including precision (P), recall (R), and mean average precision (mAP). The mAP was computed across intersection over union (IoU) thresholds ranging from 0.5 to 0.95 in increments of 0.05, following the MS-COCO evaluation protocol. Additionally, the mean average precision at IoU = 0.5 ($mAP_{0.5}$), consistent with the PASCAL VOC metric, was reported [20].

3.2 Global detection performance

Table 2 presents the quantitative comparison of YOLOv5, YOLOv7–YOLOv12 on the VIS-Day, VIS-Night, and both VIS subsets in terms of precision (P), recall (R), $mAP_{0.5}$, and mAP averaged across the ten JPEG compression levels from 50 to 5 with step of 5. The best result for each metric within a given subset is highlighted in **bold**.

For the VIS-Day subset, YOLOv12m achieves the highest precision (0.8055) and mAP (0.4625), indicating superior detection accuracy under daytime conditions. YOLOv7 attains the best recall (0.6462), while YOLOv7 and YOLOv12m provide the strongest $mAP_{0.5}$ performance, with YOLOv7 slightly higher (0.7073). Overall, YOLOv12m demonstrates the most balanced and robust performance in this subset.

In the VIS-Night subset, performance degradation is observed across all models due to challenging illumination conditions. YOLOv5m and YOLOv12m obtain the highest precision (0.4375 and 0.4372, respectively), while YOLOv10m achieves the best recall (0.2325), $mAP_{0.5}$ (0.26), and mAP (0.1517), highlighting its improved robustness in low-light scenarios.

Table 2. Detection results across daytime, nighttime, and both scenarios for VIS test set

Subset	Metric	YOLOv5	YOLOv7	YOLOv8	YOLOv9
VIS-Day	P	0.7221	0.7720	0.7657	0.7221
	R	0.6283	**0.6462**	0.6268	0.6305
	$mAP_{0.5}$	0.6736	**0.7073**	0.6905	0.6870
	mAP	0.4286	0.4496	0.4533	0.4458
VIS-Night	P	**0.4375**	0.3651	0.3980	0.4243
	R	0.2001	0.2052	0.2174	0.1823
	$mAP_{0.5}$	0.2438	0.2247	0.2492	0.2257
	mAP	0.1342	0.1235	0.1448	0.1310
VIS	P	0.7110	0.7068	0.7319	0.6646
	R	0.5276	**0.5332**	0.5285	0.5103
	$mAP_{0.5}$	0.5774	0.5808	0.5882	0.5573
	mAP	0.3705	0.3784	0.3918	0.3691

Subset	Metric	YOLOv10	YOLOv11	YOLOv12
VIS-Day	P	0.7687	0.7445	**0.8055**
	R	0.6207	0.6304	0.6349
	$mAP_{0.5}$	0.6853	0.6847	0.7048
	mAP	0.4480	0.4438	**0.4625**
VIS-Night	P	0.3891	0.3555	0.4372
	R	**0.2325**	0.1899	0.2070
	$mAP_{0.5}$	**0.2600**	0.2139	0.2462
	mAP	**0.1517**	0.1233	0.1425
VIS	P	0.7227	0.7149	**0.7754**
	R	0.5286	0.5027	0.5233
	$mAP_{0.5}$	0.5822	0.5590	**0.5908**
	mAP	0.3878	0.3699	**0.3951**

For the complete VIS set, YOLOv12m consistently outperforms other models, achieving the highest precision (0.7754), $mAP_{0.5}$ (0.5908), and mAP (0.3951). The highest recall is obtained by YOLOv7 (0.5332), though YOLOv12m remains competitive.

These results confirm that recent YOLO versions, particularly YOLOv12m and YOLOv10m, benefit from architectural and training improvements, delivering consistent performance gains over earlier versions, especially in terms of mean average precision, while maintaining competitive recall and precision across varying illumination conditions.

Table 3 presents performance comparison of YOLOv5m, YOLOv7, YOLOv8m–YOLOv12m on the LWIR-Day, LWIR-Night, and both LWIR subsets in terms of precision, recall, $mAP_{0.5}$, and mAP, averaged over ten JPEG compression levels.

Table 3. Detection results across daytime, nighttime, and both scenarios for LWIR test set

Subset	Metrics	YOLOv5	YOLOv7	YOLOv8	YOLOv9
LWIR-Day	P	0.4452	0.4424	**0.5153**	0.4398
	R	0.3951	0.4213	**0.4229**	0.3193
	$mAP_{0.5}$	0.3950	0.4213	**0.4460**	0.3603
	mAP	0.2271	0.2443	**0.2612**	0.2038
LWIR -Night	P	**0.4641**	0.4476	0.4260	0.4201
	R	0.3780	0.3815	0.4133	0.3797
	$mAP_{0.5}$	0.4009	0.3984	0.4222	0.4030
	mAP	0.2349	0.2347	0.2624	0.2495
LWIR	P	0.5243	0.5333	**0.5853**	0.4973
	R	0.4743	0.4814	**0.5057**	0.4050
	$mAP_{0.5}$	0.4887	0.5015	**0.5351**	0.4458
	mAP	0.2934	0.3062	**0.3370**	0.2726

Subset	Metrics	YOLOv10	YOLOv11	YOLOv12
LWIR-Day	P	0.4794	0.4695	0.4634
	R	0.3209	0.3303	0.3035
	$mAP_{0.5}$	0.3913	0.3683	0.3524
	mAP	0.2232	0.2043	0.1948
LWIR -Night	P	0.4407	0.4623	0.4637
	R	0.3911	0.4012	**0.4245**
	$mAP_{0.5}$	0.4239	0.4182	**0.4383**
	mAP	0.2631	0.2552	**0.2726**
LWIR	P	0.5219	0.5295	0.5384
	R	0.4247	0.4154	0.4108
	$mAP_{0.5}$	0.4897	0.4522	0.4657
	mAP	0.3014	0.2709	0.2838

For the LWIR-Day subset, YOLOv8m clearly outperforms the other models, achieving the highest precision (0.5153), recall (0.4229), $mAP_{0.5}$ (0.4460), and mAP (0.2612). This indicates that YOLOv8m provides the most accurate and reliable detections under daytime conditions.

In the LWIR-Night subset, YOLOv12m attains the best overall performance, with the highest recall (0.4245), $mAP_{0.5}$ (0.4383), and mAP (0.2726). The highest precision is obtained by YOLOv5m (0.4641), closely followed by YOLOv12m (0.4637). These results show that YOLOv12m offer improved robustness in nighttime LWIR scenarios, where thermal contrast is more pronounced.

For the complete LWIR set, YOLOv8m consistently achieves the best results across all metrics, including precision (0.5853), recall (0.5057), $mAP_{0.5}$ (0.5351), and mAP (0.3370). This confirms its strong generalization capability across both day and night thermal conditions.

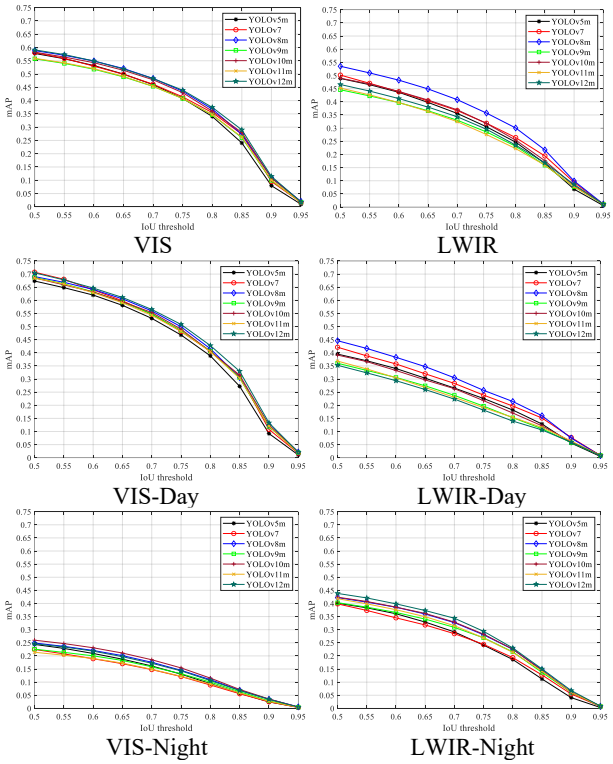


Figure 2. mAP as a function of the IoU threshold (0.5–0.95) for VIS and LWIR modalities under day and night conditions, averaged over ten JPEG compression levels (QF = 50 to 5)

The obtained results demonstrate that YOLOv8 delivers the most consistent and superior performance on LWIR data, particularly in the complete and daytime subsets, while YOLOv12m shows notable advantages under nighttime conditions.

3.3. Detection performance across IoU thresholds under JPEG compression

Figure 2 illustrates the variation in detection performance, measured in terms of mAP, as a function of the IoU, which ranges from 0.5 to 0.95 in steps of 0.05, for both visible and LWIR modalities under daytime and nighttime conditions. For each IoU threshold, the mAP is computed and then averaged over ten JPEG compression levels.

Across all configurations, the mAP decreases monotonically as the IoU threshold increases, reflecting the stricter localization requirement imposed by higher IoU values. For the VIS modality, daytime scenes consistently achieve higher mAP values than nighttime scenes at all IoU thresholds, while the performance gap becomes more pronounced at higher thresholds, indicating reduced localization accuracy under low-light conditions.

For the LWIR modality, there is a smaller gap between mAP curves in day and night scenarios. This behavior suggests that thermal imagery provides more consistent object localization under varying illumination conditions. Nevertheless, a clear performance drop is still observed.

3.4 Detection performance under varying JPEG compression levels

Figure 3 illustrates the mAP results of different YOLO object detection models under JPEG compression across QF values from 50 to 5 with step of 5, for both VIS and LWIR modalities under overall, daytime, and nighttime conditions.

Across all scenarios and YOLO models, a consistent trend is observed regarding image compression. From QF=50 to QF=20, the models remain relatively stable, demonstrating robustness to moderate compression. The mAP shows a gradual decline with a gentle slope, indicating only minor performance degradation. However, when the quality factor drops below 15, the curves exhibit a sharp "knee," with detection performance decreasing steeply as QF approaches 5. This indicates that maintaining a quality factor above 20 is essential to preserve reliable detection under compressed conditions.

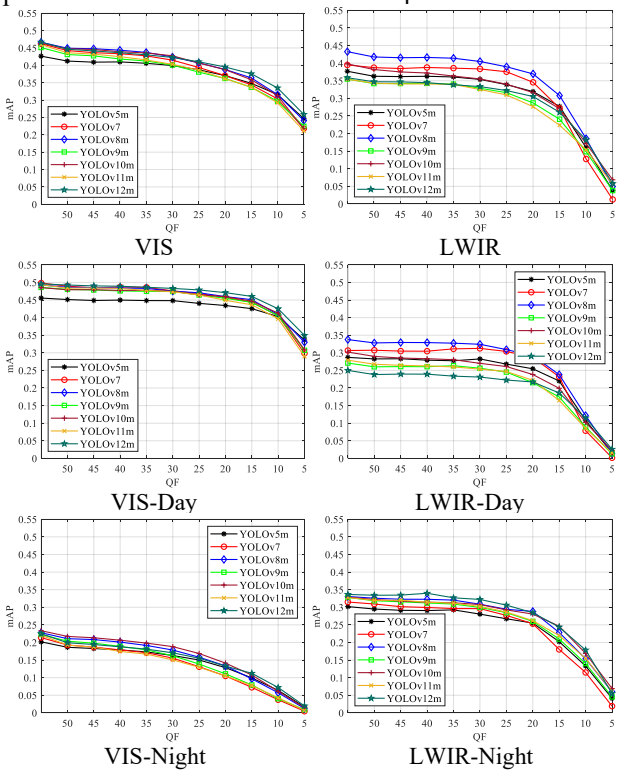


Figure 3. mAP results of YOLO models under JPEG compression across QF values for VIS and LWIR images under different lighting conditions

In comparing the models, clear differences in performance emerge. YOLOv12m and YOLOv8m consistently outperform the others across most modalities, starting from higher baselines at QF=50 and maintaining their lead as image quality decreases. The exception is VIS under nighttime conditions, where YOLOv10m achieves superior mAP values at QF levels above 20. The remaining YOLO models occupy the mid-range and frequently exchange positions. They generally outperform YOLOv5m for the visible modality and LWIR under nighttime conditions, but lag behind YOLOv8m and YOLOv12m in overall performance.

Analyzing performance across modalities reveals clear differences between VIS and LWIR images. During daytime, VIS cameras dominate, achieving mAPs near 0.50, while LWIR-Day lags behind at 0.30–0.35. This confirms that visual data provides richer texture and detail under sufficient illumination. At night, the situation reverses. VIS-Night performance drops sharply, with even the best models reaching only about 0.23 mAP, illustrating the limitations of standard cameras in low light environments. In contrast, LWIR-Night, maintains robust performance, peaking around 0.34 mAP, exceeding VIS-Night and approaching to LWIR-Day scenario. These findings indicate that thermal imaging offers consistent detection across day and night, whereas visible-light cameras are highly sensitive to lighting conditions.

3.5 Class-wise detection performance under JPEG compression

Figure 4 presents a detailed class-wise analysis of detection performance for the person, car, and bus categories. The mAP is averaged over ten JPEG compression levels.

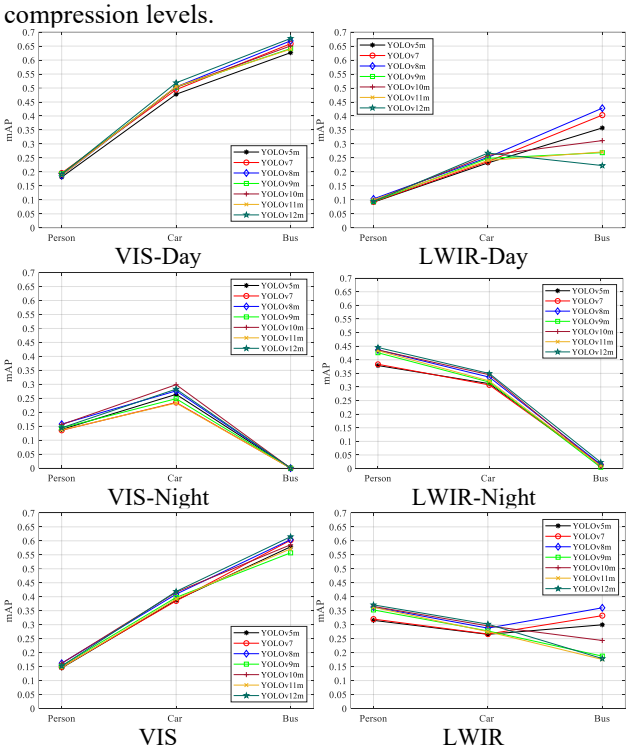


Figure 4. Class-wise detection performance in term of mAP for person, car, and bus, averaged over ten JPEG compression levels, for VIS and LWIR under daytime and nighttime conditions

In the VIS modality, daytime conditions outperform nighttime conditions for all classes, with the performance gap being most pronounced for the bus class. This indicates that YOLO detectors rely heavily on texture and color that are significantly affected by both low-light conditions and JPEG compression. The global VIS results follow the same trend as daytime conditions, confirming the dominant influence of illumination on visible-spectrum detection.

For LWIR modality, nighttime demonstrates the highest robustness for class person. The performance difference between LWIR-Day and LWIR-Night is smaller than in VIS for class car, highlighting the reduced sensitivity of thermal imagery to illumination changes.

Across all scenarios, the car class consistently achieves a stable mAP, reflecting its stable visual and thermal signatures, which make it less sensitive to compression artifacts. The person class exhibits moderate performance, with noticeable differences between day and night conditions, particularly in the LWIR modality, where reduced illumination and compression jointly degrade fine-grained features. The bus class shows more variable behavior, which can be attributed to its lower number of instances within nighttime scenario.

4. CONCLUSION

This study provides a comprehensive evaluation of JPEG compression effects on road scene object detection using YOLO models from YOLOv5m to YOLOv12m across visible and LWIR imagery. The results demonstrate that recent YOLO architectures, particularly YOLOv12m and YOLOv8m, benefit from architectural and training improvements, achieving higher precision, recall, and mAP under both daytime and nighttime conditions. VIS detection performance is highly sensitive to illumination and compression, with daytime scenes outperforming nighttime, whereas LWIR offers more consistent detection across different lighting conditions. Compression impacts all models, with a gradual mAP decline for QF values above 20 and a steep drop below QF=20, emphasizing the need to maintain moderate image quality. The analysis also reveals modality-specific trends, VIS provides richer texture under sufficient light, while LWIR ensures robust detection at night. These findings underscore the importance of considering both compression level and imaging modality when deploying real time road scene object detection systems, and they provide practical guidelines for maintaining reliable performance in bandwidth limited scenarios. Future work may explore alternative compression schemes, additional detection architectures, and adaptive training strategies on compressed data to further improve detection performance.

References

[1] ADAM M.A.A., TAPAMO, J.R.: *Survey on image-based vehicle detection methods,* World Electric Vehicle Journal 16(6), 303 (2025).

[2] GANDOR, T., NALEPA, J.: First gradually, then suddenly: understanding the impact of image compression on object detection using deep learning, Sensors 22(3), 1104 (2022).

[3] BAKIR, H., Bakir, R.: Evaluating the robustness of YOLO object detection algorithm in terms of detecting objects in noisy environment, Journal of Scientific Reports-A 054, 1–25 (2023).

[4] JOCHER, G., AYUSH, C., JING, Q.: *Ultralytics YOLOv8,* (2023). [Online]. Available: https://github.com/ ultralytics/ ultralytics. Accessed 2 Aug 2025

[5] Terven, J., CÓRDOVA-ESPARZA, D.-M., ROMERO-GONZÁLEZ, J.-A.: *A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS,* Machine Learning and Knowledge Extraction 5(4), 1680–1716 (2023).

[6] Glenn, J., JING, Q.: *Ultralytics YOLOv11,* (2024). [Online]. Available: https://github.com/ultralytics/ ultralytics. Accessed 2 Aug 2025

[7] DODGE, S., KARAM, L.: *Understanding how image quality affects deep neural networks,* In: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6, Lisbon, Portugal, (2016).

[8] AQQA, M., MANTINI, P., SHAH, S.K.: *Understanding how video quality affects object detection algorithms,* In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Application, pp. 96–104, Prague, Czech Republic, (2019).

[9] BAYERL, A., KEGLEVIC, M., WÖDLINGER, M.G., SABLATNIG, R.: *Impact of learned domain specific compression on satellite image object classification*, In: Proceedings of the 26th Computer Vision Winter Workshop (CVWW 2023); Austria, (2023).

[10] Tanaka, T., HAREL, A., BAJIĆ, I.V.: *Does video compression impact tracking accuracy?,* In: 2022 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1517–1521, Austin, TX, USA, (2022).

[11] ADLI, T., BUJAKOVIĆ, D.M., BONDŽULIĆ, B.P., Laidouni, M.Z., ANDRIĆ, M.S.: *Robustness of YOLO models for object detection in remote sensing images,* Journal of Electrical Engineering 76(5), 429–442 (2025).

[12] Jocher, G.: *YOLOv5 by Ultralytics,* (2020). [Online]. Available: https://github.com/ultralytics/yolov5. Accessed 2 Aug 2025

[13] WANG, C.Y., BOCHKOVSKIY, A., LIAO, H.Y.M.: *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475, Vancouver, BC, Canada, (2023).

[14] WANG, C.Y., YEH, I.H., LIAO, H.Y.M.: YOLOv9: *Learning what you want to learn using programmable gradient information,* In: European Conference on Computer Vision, pp. 1–21, Milan,

Italy, (2024).

[15] CHEN, H., et al.: YOLOv10: *Real-time end-to-end object detection,* In: Advances in Neural Information Processing Systems 37 (NeurIPS 2024), pp. 107984–108011, San Diego, California, USA, (2024).

[16] Tian, Y., YE, Q., DOERMANN, D.: YOLOv12: *Attention-centric real-time object detectors,* arXiv (2025). arXiv:2502.12524.

[17] LIU, J., FAN, X., HUANG, Z., WU, G., LIU, R., ZHONG, W., LUO, Z.: Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection, In: Proceedings of the IEEE/CVF Conference on Com-puter Vision and Pattern Recognition, pp. 5802–5811, New Orleans, LA, USA, (2022).

[18] Hudson, G., LÉGER, A., NISS, B., SEBESTYÉN, I., VAABEN, J.: *JPEG-1 standard 25 years: past, present, and future reasons for a success,* Journal of Electronic Imaging 27(4), 040901 (2018).

[19] LIN, T.-Y., MAIRE, M., BELONGIE, S., et al.: *Microsoft COCO: Common objects in con-text,* In: European Conference on Computer Vision, pp. 740–755, Zurich, Switzerland (2014).

[20] Padilla, R., NETTO, S.L., DA SILVA, E.A.B.: *A survey on performance metrics for object-detection algorithms,* In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242, Niteroi, Brazil, (2020).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AUTONOMOUS ROBOTIC SYSTEMS FOR 3D MOBILE MAPPING IN UNDERGROUND MINING

NENAD VUŠOVIĆ
University of Belgrade, Technical Faculty in Bor, Bor, nvusovic@tfbor.bg.ac.rs

MILICA VLAHOVIĆ
University of Belgrade, Institute of Chemistry, Technology and Metallurgy-National Institute of the Republic of Serbia, Belgrade, m.vlahovic@ihtm.bg.ac.rs, mvlahovic@tmf.bg.ac.rs

Abstract: *The ongoing digital transformation of the mining industry has led to the increasing adoption of advanced geospatial technologies and autonomous robotic systems within the smart mining paradigm. This study examines the application and capabilities of autonomous 3D mobile mapping in underground mining environments, with particular emphasis on robotic platforms equipped with LiDAR and complementary sensor technologies. Conventional surveying methods in underground conditions are constrained by limited visibility, confined geometries, and elevated safety risks, often resulting in time-intensive data acquisition and incomplete spatial information. In contrast, autonomous mobile mapping systems enable rapid, high-resolution data collection in the form of dense point clouds, significantly enhancing both the accuracy and efficiency of spatial data acquisition. The study analyzes key enabling technologies, including terrestrial laser scanning (TLS), LiDAR-based systems, and autonomous robotic platforms, as well as their integration within geographic information systems (GIS) for advanced spatial analysis and visualization. Special consideration is given to the deployment of robotic systems in hazardous and inaccessible zones, where they minimize human exposure while enabling continuous and reliable monitoring. The findings demonstrate that the implementation of autonomous robotic mapping systems provides substantial improvements in data quality, operational efficiency, and safety. Furthermore, their integration supports the generation of 4D spatiotemporal models, facilitating improved decision-making and long-term planning in underground mining operations.*

Keywords: *LiDAR, terrestrial laser scanning.*

1. INTRODUCTION

The mining industry is undergoing an intensive digital transformation driven by the increasing demand for higher productivity, improved operational safety, sustainable resource management, and reduced production costs [1]. Within this context, the concept of Smart Mining has emerged as a comprehensive framework integrating advanced geospatial technologies, artificial intelligence, industrial communication networks, cloud computing, and autonomous robotic systems into a unified digital environment for mine management [2,3]. Underground mining represents one of the most demanding engineering environments due to confined spaces, poor visibility, dust, humidity, unstable geological conditions, and the absence of Global Positioning System (GPS) signals. These limitations significantly complicate conventional surveying operations and reduce the efficiency of spatial data acquisition. At the same time, underground environments expose personnel to numerous hazards, including rockfalls, toxic gases, ventilation problems, flooding, and mechanical risks associated with mining equipment [2,4].

Reliable spatial information is fundamental to virtually every stage of underground mining, including mine development, excavation control, deformation monitoring, ventilation planning, volume estimation, and long-term production management [5]. Traditionally, these tasks have relied on total stations, leveling networks, and conventional mine surveying techniques. Although these methods provide satisfactory accuracy, they require considerable time in the field and usually acquire only a limited number of discrete measurement points.

Recent advances in terrestrial laser scanning (TLS), LiDAR technology, simultaneous localization and mapping (SLAM), and autonomous robotic navigation have fundamentally changed the methodology of underground spatial data acquisition. Modern autonomous mobile mapping systems rapidly collect millions of three-dimensional points, producing dense point clouds that accurately represent the geometry of underground excavations while significantly reducing the need for direct human intervention in hazardous environments [3]. While TLS provides high-accuracy static data acquisition, autonomous mobile mapping extends these capabilities through continuous SLAM-based data collection.

Unlike conventional surveying, autonomous robotic mapping systems combine multiple sensors, including LiDAR scanners, inertial measurement units (IMUs), RGB cameras, thermal cameras, odometers, and environmental sensors, into a single integrated platform capable of autonomous navigation and real-time spatial data acquisition. The collected datasets are processed into

Corresponding author: M. Vlahović;
E-mail: m.vlahovic@ihtm.bg.ac.rs,
DOI: 10.5937/OTEH26061V

high-resolution three-dimensional models suitable for inspection, deformation analysis, infrastructure monitoring, volume calculations, and digital mine management.

Beyond underground mining, these technologies have attracted increasing attention in numerous safety-critical applications. Autonomous robotic mapping systems are now employed for the inspection of transportation tunnels, underground infrastructure, military facilities, emergency response operations, disaster assessment, border-security tunnels, and other environments where direct human access is difficult or dangerous. Their capability to rapidly generate accurate digital models while minimizing personnel exposure makes them particularly valuable for both civilian and defense-related engineering applications.

The objective of this paper is to present the principles, technological components, and practical applications of autonomous robotic systems for 3D mobile mapping in underground mining. Particular emphasis is placed on LiDAR-based mobile mapping technologies, robotic platforms, integration with Geographic Information Systems (GIS), and their contribution to improving operational efficiency, safety, and decision-making through the generation of high-resolution three-dimensional and four-dimensional spatial models.

2. CONVENTIONAL SURVEYING IN UNDERGROUND MINES

Accurate spatial information forms the basis for planning, construction, operation, and monitoring throughout the entire life cycle of an underground mine. Mine surveying supports exploration, excavation planning, construction of underground structures, deformation monitoring, production control, reserve estimation, and preparation of mine documentation [5]. Consequently, the quality of surveying data directly influences operational efficiency, production planning, and occupational safety.

Traditional underground surveying relies primarily on mine control networks established by polygon traverses and leveling systems connected to the surface reference network [5]. Depending on the mining stage and required accuracy, surveying activities include staking out underground excavations, monitoring shaft geometry, controlling excavation profiles, deformation measurements, determination of mined volumes, and preparation of mine plans.

Despite continuous improvements in surveying instruments, underground mining remains an extremely challenging working environment. Underground excavations are characterized by narrow passages, limited visibility, high humidity, dust, unstable rock masses, and the presence of hazardous gases. Furthermore, the absence of GNSS signals prevents the direct application of satellite-based positioning technologies, while magnetic disturbances may reduce the performance of certain navigation sensors [2,5].

Conventional surveying is predominantly performed using total stations, laser distance meters, leveling instruments, and prism-based measurements. These techniques provide high measurement accuracy but require extensive field operations, careful instrument setup, repeated occupation of survey stations, and direct access to the measured object. Consequently, spatial information is collected only at selected points or along predefined cross-sections, making the resulting representation only a partial description of complex underground geometries [5].

For example, excavation profiles are traditionally surveyed using direct measurements, arc intersection methods, or polar measurements performed with total stations. Although these procedures satisfy engineering requirements for many routine surveying tasks, they remain relatively slow and labor-intensive, especially when large underground areas must be documented [6].

Periodic inspection of underground structures represents another important surveying task. Monitoring mine shafts, haulage drifts, ventilation shafts, and transport galleries traditionally requires repeated measurements of predefined control points installed within concrete linings or surrounding rock masses.

Deformation monitoring is usually performed through convergence measurements and geodetic observations of permanently installed benchmarks (Figure 1), enabling the determination of displacement vectors and assessment of structural stability [7].

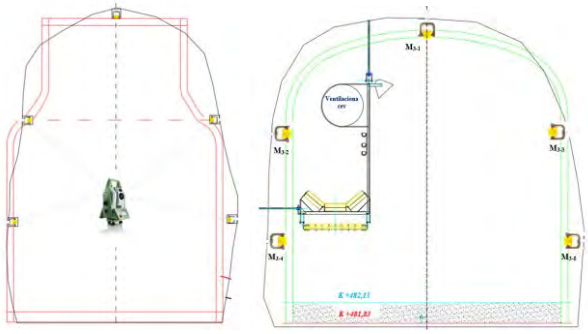


Figure 1. Conventional deformation monitoring using control benchmarks [7]

A particularly demanding surveying task is the inspection of vertical mine shafts. Conventional inspection procedures require surveyors to work inside confined vertical structures using suspended platforms or scaffolding while measuring only selected cross-sections (Figure 2). Such procedures are time-consuming, technically demanding, and expose personnel to considerable safety risks. Moreover, only limited geometric information can be obtained, making comprehensive structural assessment difficult [5].

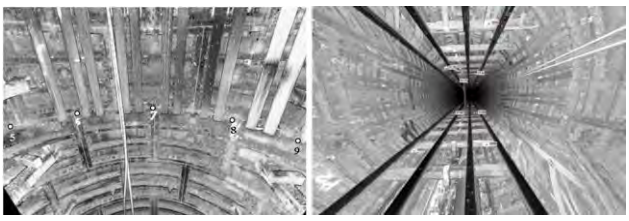


Figure 2. Conventional inspection of mine shafts [5].

Traditional methods are also employed for estimating excavated ore volumes. Volume calculations are generally based on measurements of excavation cross-sections or selected characteristic points acquired by total stations. Since only discrete points are measured, the generated two-dimensional models represent an approximation of the actual excavation geometry, often resulting in reduced accuracy of volume estimation and production assessment [5].

The principal limitations of conventional surveying methods include prolonged fieldwork, incomplete spatial coverage, dependence on direct human access, and relatively low efficiency when documenting large or geometrically complex underground environments. In hazardous locations, these limitations become even more pronounced, as the safety of surveying personnel must always take precedence over measurement efficiency.

These shortcomings have accelerated the adoption of autonomous mobile mapping technologies capable of rapidly acquiring complete three-dimensional spatial information while significantly reducing the need for human presence in hazardous underground areas.

3. AUTONOMOUS ROBOTIC SYSTEMS FOR 3D MOBILE MAPPING

The rapid development of digital technologies has fundamentally transformed spatial data acquisition in underground mining. Advances in robotics, laser scanning, artificial intelligence, computer vision, wireless communication, and geospatial information systems have enabled the transition from conventional surveying toward fully autonomous mobile mapping systems. These technologies represent one of the key components of the Smart Mining concept, providing continuous acquisition of high-resolution spatial information while significantly reducing human exposure to hazardous environments [2,3].

Unlike traditional surveying methods, autonomous robotic mapping systems continuously collect spatial information while moving through underground excavations. Instead of measuring selected points or cross-sections, they generate dense three-dimensional point clouds describing the complete geometry of underground structures with high spatial resolution. The resulting digital models enable detailed visualization, quantitative analysis, deformation monitoring, and long-term documentation of underground infrastructure.

A typical autonomous mapping platform integrates several complementary technologies, including LiDAR sensors, inertial measurement units (IMUs), simultaneous localization and mapping (SLAM) algorithms, digital cameras, odometry sensors, and onboard processing units. The combination of these technologies enables accurate localization even in environments where GPS positioning is unavailable, which is one of the principal challenges in underground applications.

The collected spatial information is commonly integrated into Geographic Information Systems (GIS), where it becomes available for visualization, spatial analysis, inspection planning, infrastructure management, and decision support. Continuous data acquisition at different time intervals also enables the creation of four-dimensional (4D) models that describe not only the geometry of underground excavations but also their temporal evolution, making it possible to monitor structural changes and deformation processes throughout the operational life of the mine [3].

3.1 LiDAR-Based Mobile Mapping

Light Detection and Ranging (LiDAR) technology has become one of the most important sensing techniques for autonomous underground mapping. LiDAR sensors determine object positions by measuring the travel time of emitted laser pulses reflected from surrounding surfaces, producing dense three-dimensional point clouds with high geometric accuracy.

Compared with conventional surveying techniques, LiDAR systems offer several important advantages. They rapidly collect millions of measurement points, operate independently of ambient lighting conditions, and accurately capture irregular geological structures, tunnel profiles, mining chambers, shafts, and excavation faces. Because measurements are performed continuously during platform movement, the entire underground environment can be documented much more efficiently than with traditional point-by-point surveying [3].

The resulting point clouds form the basis for numerous engineering analyses, including cross-sectional measurements (Figure 3), deformation assessment, excavation volume calculation, infrastructure inspection, geological mapping, and digital mine documentation.

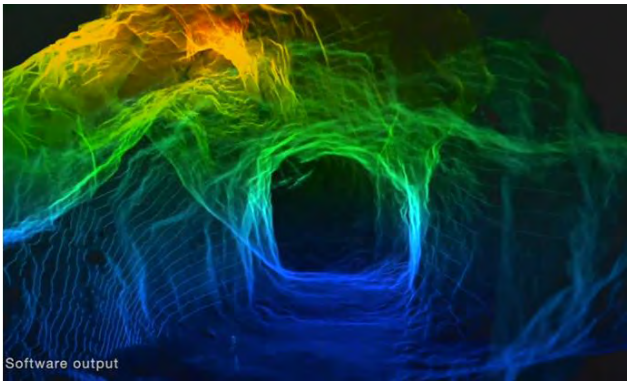


Figure 3. Mobile scanning of underground tunnels with LiDAR and point clouds software output [5]

3.2 SLAM-Based Navigation

Since satellite positioning systems cannot operate underground, autonomous mapping systems rely on Simultaneous Localization and Mapping (SLAM) algorithms. SLAM enables a robotic platform to determine its position while simultaneously constructing a map of an unknown environment using sensor observations.

LiDAR-based SLAM continuously matches newly acquired point clouds with previously recorded spatial information, allowing the robot to estimate its trajectory

with high accuracy even in long underground tunnels and complex mining areas (Figure 4). Modern SLAM algorithms combine laser measurements with data obtained from inertial sensors, wheel encoders, and visual cameras, improving positioning robustness under challenging operating conditions.

The application of SLAM significantly reduces cumulative positioning errors and enables autonomous navigation through extensive underground environments without requiring external positioning infrastructure [8,9].

Figure 4. LiDAR-based SLAM for mapping in high-risk mining areas [5]

Handheld mobile mapping systems represent another important category. These portable LiDAR devices enable rapid acquisition of spatial information while operators simply walk through underground excavations (Figure 5). Although handheld systems still require human presence, they dramatically reduce field time compared with conventional surveying techniques and generate considerably more comprehensive spatial datasets.

Figure 5. Handheld GeoSLAM LiDAR system for autonomous 3D mobile mapping [10]

3.3 Robotic Mobile Platforms

Recent years have witnessed the development of various robotic platforms specifically designed for underground mapping applications. Depending on operational requirements, autonomous systems may be classified as ground robotic vehicles (UGVs), aerial robotic platforms (UAVs), or hybrid systems combining multiple sensing technologies.

Ground robotic vehicles equipped with LiDAR scanners can autonomously inspect underground drifts, transport galleries, ventilation tunnels, shafts, and production areas while continuously collecting three-dimensional spatial information. Their ability to operate in hazardous or inaccessible locations substantially reduces the need for personnel exposure during routine inspection and surveying operations [10]. A scheme of an autonomous robotic mapping platform is given in Figure 6.

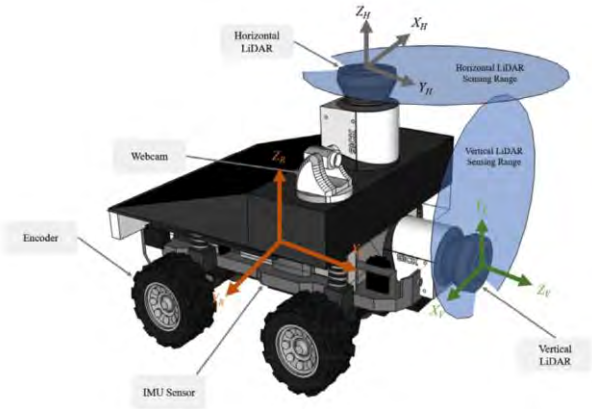


Figure 6. Conceptual configuration of an autonomous robotic mapping platform [11]

Unmanned aerial vehicles (UAVs) are increasingly being adopted for mapping large underground chambers, stopes, and vertical shafts where direct access is difficult or impossible. Equipped with lightweight LiDAR sensors or optical cameras, UAVs can safely inspect areas that would otherwise require complex and potentially hazardous surveying procedures [12].

Figure 7 presents the mobile mapping using a UAV equipped with LiDAR and the resulting point cloud.

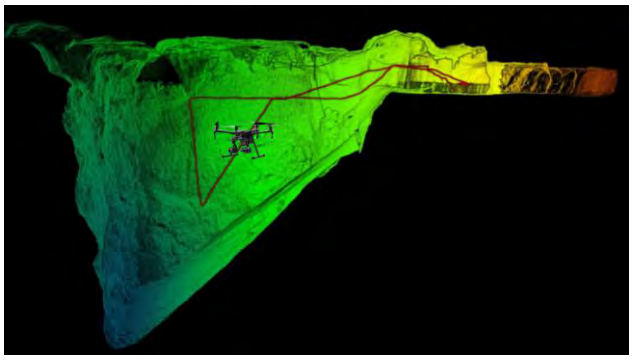


Figure 7. 3D models of the excavated space in the underground mine captured by a drone [13]

3.4 Advantages of Autonomous Robotic Mapping

The implementation of autonomous robotic mapping systems provides significant improvements compared with conventional surveying approaches. The most important advantages include:

- rapid acquisition of complete three-dimensional spatial information;
- generation of dense, high-resolution point clouds;
- reduced field survey time;

- improved operator safety through reduced exposure to hazardous environments;
- high repeatability of measurements suitable for long-term monitoring;
- direct integration with GIS and digital mine management systems;
- supports for the development of digital twins and four-dimensional (4D) spatiotemporal models;
- improved support for engineering analysis, production planning, and infrastructure maintenance.

These capabilities make autonomous robotic systems particularly suitable for continuous monitoring of underground infrastructure, where both measurement quality and personnel safety are of primary importance.

4. APPLICATIONS IN UNDERGROUND MINING AND SAFETY-CRITICAL INFRASTRUCTURE

Autonomous robotic mapping systems have become increasingly important in underground mining because they provide rapid and reliable spatial information while reducing personnel exposure to hazardous environments. Their application extends across the entire mining cycle, from mine development and infrastructure inspection to deformation monitoring and production control.

One of the most significant applications is the inspection of underground openings such as shafts, haulages, ventilation drifts, transport galleries, and mining chambers. Conventional inspection procedures require direct human access and often provide only partial geometric information. In contrast, autonomous mobile mapping systems acquire complete three-dimensional representations of underground structures, enabling detailed visualization and quantitative assessment of their condition [5].

Periodic scanning of underground excavations also enables deformation monitoring through comparison of point clouds acquired at different epochs. The generated datasets can be integrated within GIS environments to create four-dimensional (4D) models that describe both geometry and temporal changes, supporting stability assessment and long-term infrastructure management.

Another important application is the determination of excavated ore volumes. Dense point clouds generated by LiDAR-based mapping systems provide significantly more complete geometric information than conventional profile measurements, improving the accuracy of volume calculations and production analysis [14].

Because autonomous systems can operate in areas with limited accessibility, they are particularly valuable for inspection of hazardous zones affected by rock instability, poor ventilation, flooding, or other operational risks. The reduction of direct human exposure represents one of the principal advantages of robotic deployment in underground environments.

Although underground mining provides the primary application context considered in this study, the same technological principles are increasingly employed in other safety-critical environments characterized by confined geometry, limited visibility, and difficult access. Examples include transportation tunnels, underground utility corridors, storage caverns, emergency shelters, and critical infrastructure facilities.

Of particular relevance for defense and security applications is the use of autonomous robotic mapping systems for inspection and reconnaissance of underground structures where direct human entry may be undesirable or unsafe. Robotic platforms equipped with LiDAR and visual sensors can rapidly generate accurate three-dimensional models of tunnels, galleries, shelters, and other enclosed spaces while maintaining operator standoff distance. Such capabilities support situational awareness, route assessment, infrastructure inspection, and post-event damage evaluation.

Similarly, during emergency response and rescue operations, autonomous mapping systems can provide rapid spatial information in collapsed or unstable underground environments, helping responders identify accessible routes, obstacles, and areas requiring immediate attention.

The common denominator across mining, civil, and defense-related applications is the ability of autonomous robotic systems to acquire high-resolution spatial data in environments that are hazardous, inaccessible, or operationally constrained.

This combination of safety, speed, and geometric completeness is the main reason for the growing adoption of robotic mobile mapping technologies in modern underground operations.

5. DISCUSSION

The transition from conventional surveying toward autonomous robotic mapping represents a significant technological change in underground spatial data acquisition. Traditional surveying methods remain essential for establishing mine control networks and achieving high local accuracy; however, they are inherently limited by the need for direct human access and by the relatively sparse spatial information that can be collected during routine operations [5].

Autonomous mobile mapping systems address many of these limitations by acquiring dense three-dimensional datasets during continuous platform movement. The resulting point clouds provide a much more complete representation of underground geometry, enabling analyses that are difficult to achieve using discrete measurements alone. In addition, repeated scanning campaigns support temporal monitoring and the generation of 4D models, which are increasingly important for stability assessment and infrastructure management [3]. From an operational perspective, the most important benefit is the reduction of personnel exposure in hazardous underground environments.

Inspection of shafts, unstable excavations, flooded areas, or poorly ventilated zones can often be performed remotely by robotic platforms, improving safety while

maintaining the availability of high-quality spatial information.

The broader applicability of these systems is also noteworthy. The same combination of LiDAR sensing, SLAM-based navigation, robotic mobility, and GIS integration can be employed in mining, transportation tunnels, critical infrastructure inspection, emergency response, and defense-related underground operations. This technological convergence suggests that advances achieved in one domain can be transferred efficiently to other sectors with similar operational constraints [15,16].

6. CONCLUSIONS

This paper examined the application of autonomous robotic systems for 3D mobile mapping in underground mining environments. The analysis demonstrated that LiDAR-based mobile mapping, SLAM navigation, and, where appropriate, TLS implemented on handheld systems and autonomous robotic platforms, enable the rapid acquisition of dense three-dimensional spatial information even in environments where GNSS positioning is unavailable. Compared with conventional surveying methods, autonomous systems provide more complete geometric coverage, shorter field acquisition times, improved repeatability, and significantly enhanced operator safety. Their integration with GIS platforms enables advanced visualization, spatial analysis, deformation monitoring, and the generation of four-dimensional spatiotemporal models that support planning and decision-making throughout the mining life cycle.

The study also highlighted that the capabilities developed for underground mining are directly applicable to other safety-critical environments, including transportation tunnels, underground infrastructure, emergency-response scenarios, and defense-related underground facilities. The ability to obtain accurate spatial information while minimizing human exposure is the key factor driving the increasing adoption of autonomous robotic mapping technologies across multiple engineering and security domains.

Future developments are expected to focus on greater autonomy, improved multi-sensor fusion, real-time data processing, and tighter integration with digital-twin and intelligent decision-support systems, further strengthening the role of autonomous robotic mapping in underground operations [17].

Acknowledgement

The authors would like to thank the Ministry of Science, Technological Development and Innovations of the Republic of Serbia for support (Grant Nos: 451-03-34/2026-03/200131 and 451-03-33/2026-03/200026).

References

[1] CUI,S., BAO,J., HU,D., YUAN,X., ZHANG,K., YIN,Y.,WANG,M., ZHU,C.: *Research status and development trends of SLAM technology in autonomous mining field*, Journal of Mine Automation, 50(10) (2024) 38-52.

[2] VUŠOVIĆ,N.: *Mine Design (Authorized Lectures)*, University of Belgrade, Technical Faculty in Bor, Bor, 2022.

[3] VUŠOVIĆ,N.: *Geoinformatics*, University of Belgrade, Technical Faculty in Bor, Bor, 2023.

[4] MILIĆ,V., MILIĆEVIĆ,Ž.: *Fundamentals of Mineral Resource Exploitation*, University of Belgrade, Technical Faculty in Bor, Bor, 2005.

[5] VUŠOVIĆ,N.: *Historical overview and concept of modern geodesy and mining measurements*, University of Belgrade, Technical Faculty in Bor, Bor, 2004.

[6] GANIĆ,A., MILUTINOVIĆ,A., TOKALIĆ,R., OGNJANOVIĆ,S.: *Measuring methods for cross sections of underground mine chambers*, Underground mining engineering, 19 (2011) 101-108.

[7] VUŠOVIĆ,N.: *Project on observation of the stability of tunnel lining with chambers at the Suvo Vrelo open-pit of Titan-Kosjerić cement plant*, University of Belgrade, Technical Faculty in Bor, Bor, 2021.

[8] Zhang,Y., Shi,P., Li,J.: *3D LiDAR SLAM: A survey*, The Photogrammetric Record, 39 (2024) 457-517.

[9] EBADI,K., BERNREITER,L., BIGGIE,H., et al.: *Present and Future of SLAM in Extreme Environments: The DARPA SubT Challenge*, IEEE Transactions on Robotics, 40 (2022) 936-959.

[10] TATSCH,C., BREDU,J.,A.,JNR, COVELL,D., TULU,I.,B., GU,Y.: *RHINO: AN AUTONOMOUS ROBOT FOR MAPPING UNDERGROUND MINE ENVIRONMENTS*, COMPUTER SCIENCE ROBOTICS, 2023.

[11] RYAN,P.,M., BADYAL,A., SZE,S., et al.: *Safety assurance challenges for autonomous drones in Underground Mining Environments.* In: Huda,

[12] M.N., Wang, M. and Kalganova, T., (eds.) Towards Autonomous Robotic Systems, TAROS 2024. 25th Towards Autonomous Robotic Systems Conference, 21-23 Aug 2024. Springer, pp. 169-181.

[13] HEONMOO,K., YOSOON,C.: Location estimation of autonomous driving robot and 3D tunnel mapping in underground mines using pattern matched LiDAR sequential images, International Journal of Mining Science and Technology, 31 (2021) 779-788.

[14] https://geoslam.com/technology/ (Accessed 1 July 2026).

[15] INOSTROZA,F., PARRA-TSUNEKAWA,I., RUIZ-DEL-SOLAR,J.: *Robust Localization for Underground Mining Vehicles: An Application in a Room and Pillar Mine*, Sensors, 23(19) (2023) 8059. https://doi.org/10.3390/s23198059

[16] SAID,K.,O., SHARMA,P., AGARWAL,S., BHATTACHARYYA,S., KUMAR,A.,R.: *Advances in autonomous navigation for underground mining operations: Driven by camera, LiDAR, and radar-based SLAM*, Journal of Industrial Safety, Article in press, https://doi.org/10.1016/j.jinse.2026.05.001.

[17] LI J., ZHAN K. Intelligent Mining Technology for an Underground Metal Mine Based on Unmanned Equipment. Engineering (2018) Vol. 4, Issue 3: 381-391.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EDGE AI DRONE CLASSIFICATION IN RADIO FREQUENCY DOMAIN

VASILIJA JOKSIMOVIĆ
Military Technical Institute, Belgrade, vasilija.joksimovic@mod.gov.rs

DARKO MIKANOVIĆ
Military Technical Institute, Belgrade, darko.mikanovic@mod.gov.rs

RADIANA BRUSIN
Military Technical Institute, Belgrade, himerra@gmail.com

DRAGANA BOJIĆ,
Military Technical Institute, Belgrade, dragana.bojic@mod.gov.rs

Abstract: *Unmanned aerial systems (UAS) are increasingly present in modern operations, creating the need for efficient and reliable anti-drone detection systems. Artificial intelligence (AI), particularly real-time detection algorithms such as YOLO, provides fast and accurate analysis capabilities, improving situational awareness and response efficiency. This paper presents a hardware-oriented approach for detecting radio frequency (RF) communication between drones and their controllers. Since raw RF signals do not provide sufficient discriminative features, they are transformed into spectrogram representations, enabling feature extraction and subsequent classification [1]. The proposed system integrates RF signal acquisition, preprocessing, and YOLO-based detection, with the complete inference process deployed on an Edge AI platform. The solution was implemented and evaluated on the NVIDIA Jetson Orin device, demonstrating the feasibility of real-time drone communication detection in distributed C-UAS applications. Experimental results, including continuous execution and multi-process scenarios, confirmed stable operation with efficient resource utilization, while the integration of energy-based detection further improved computational efficiency by reducing unnecessary inference operations. These results highlight the potential of combining RF spectrogram analysis, deep learning, and Edge AI optimization for autonomous, low-latency, and energy-efficient C-UAS systems.*

Keywords: *Edge AI, C-UAS, RF detection, YOLO, spectrum analysis*

1. INTRODUCTION

This paper presents a hardware implementation of a convolutional neural network (CNN) based on the YOLOv8 algorithm, designed for counter-unmanned aerial systems (C-UAS) applications. The proposed approach focuses on the detection and classification of radio frequency (RF) communication between drones and their control stations. Since direct observation of RF signals does not provide sufficient discriminatory features for classification, the signals are transformed into spectrogram representations, enabling the extraction of relevant patterns for further analysis [2, 3].

Based on the acquired and preprocessed RF data, a dataset was constructed to train the YOLOv8 model. The network output corresponds to predefined classes, including scenarios with a single drone where the drone type can be identified as well as multi-drone scenarios, where only the number of detected drones is determined. YOLO-based classification has demonstrated strong potential in distinguishing UAV communication patterns from spectrogram data [4, 5].

The main contribution of this work is the implementation and evaluation of the proposed model on an Edge AI platform, specifically the NVIDIA Jetson Orin device. Given the time-sensitive requirements of C-UAS systems, the study emphasizes real-time processing capabilities and computational efficiency. In particular, the work investigates the feasibility of deploying deep learning models directly at the edge, reducing reliance on cloud infrastructure and minimizing latency in operational scenarios. Additionally, the analysis considers hardware constraints such as power consumption, memory usage, and processing throughput, which are critical for real-world deployment of embedded AI systems [6,7].

The model was evaluated through two processing approaches: simulation of a real-time execution on dedicated hardware and offline processing using a pre-collected dataset. This dual evaluation enables a direct comparison of performance and provides insight into the practical applicability of Edge AI solutions in operational environments.

Although the results demonstrate the feasibility of deploying such models on edge devices, the study is limited by the relatively small training dataset, which affects classification accuracy and generalization.

Corresponding author: V. Joksimović;
E-mail: vasilija.joksimovic@mod.gov.rs
DOI: 10.5937/OTEH26062J

The paper is organized as follows: Section 2 presents the Edge AI implementation and describes the motivation for deploying deep learning-based classification algorithms on embedded hardware platforms. Section 3 describes the research methodology, including RF signal acquisition, spectrogram generation, dataset preparation, and the experimental setup used for evaluation. Section 4 presents the obtained results and provides an analysis of inference performance, hardware utilization, and the impact of energy-based detection on system efficiency. Finally, Section 5 summarizes the main findings and concludes the study, highlighting the potential of Edge AI-based solutions for C-UAS applications.

2. EDGE AI IMPLEMENTATION

The concept of Edge AI enables data processing directly at the source of generation on edge devices, reducing latency, enhancing response times, and minimizing the need for continuous data transmission to centralized servers. This capability is especially beneficial for time-sensitive applications, such as drone classification, where rapid decision-making is crucial. In the realm of security and defense, particularly in the context of drone classification, these advantages are essential. Real-time classification and decision-making are often necessary, and reliance on cloud infrastructure can lead to unacceptable delays or communication bottlenecks. Moreover, Edge AI enhances system resilience and operational autonomy by reducing dependence on persistent network connectivity, which is especially valuable in remote or contested environments where cloud access may be limited or unreliable. These features make Edge AI a compelling choice for implementing intelligent, distributed anti-drone defense systems. Given the time-sensitive nature of counter-drone operations the primary objective of this study was to explore the application of the NVIDIA Jetson Orin, a dedicated edge GPU computer, and to evaluate its processing efficiency in comparison to conventional approaches [6, 8].

3. METHODOLOGY

The research methodology is illustrated in segments, as shown in Figure 1. Each segment will be described in greater detail in the following sections.

The first segment involves the acquisition of raw RF communication signals from drones' communication, followed by their processing, with spectrograms generated as the output of this stage.

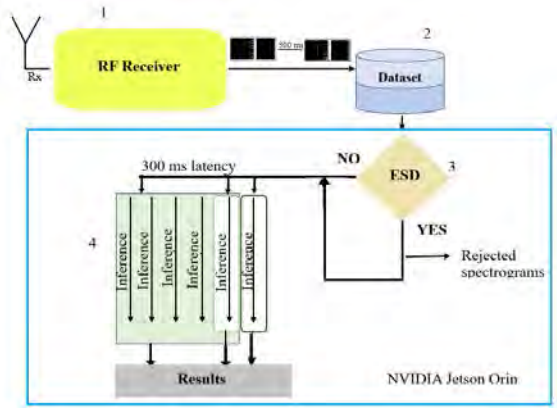


Figure 1. Methodology (1 – RF receiver, 2 – database for offline analysis and generating a dataset for model training, 3 – energy detection on spectrograms, 4 – Edge AI hardware Inference)

The RF receiver generates the spectrograms across four frequency bands: 400 MHz, 800 MHz, 2.4 GHz, and 5.8 GHz. Signal acquisition was conducted continuously, with the data captured in the form of *.txt files. Within each frequency band, three files were generated per second, each with dimensions of 2048×1000. The discrete Fourier transform length (N_DFT) was set to 2048, and the number of frames per file was 1000. Following signal acquisition, subsequent morphological operations were applied to the resulting spectrograms. These processed spectrograms were transferred into the black and white spectrograms. The total processing time, which includes both the acquisition and the computation phases is approximately 300 ms for 1000 frames. Representative examples of these generated spectrogram images are provided in Figure 2. [9, 10].

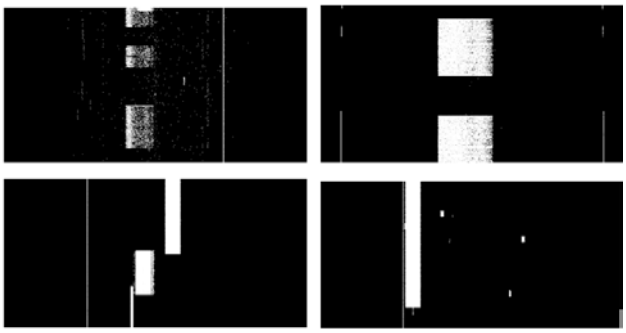


Figure 2. Generated black and white spectrograms

The second segment shown in the figure, represents a non-publicly available (VTI_RFDroneSET_USRP) database created using the outputs from the first segment. This database contains spectrograms intended for either the training or testing of existing or new models. The data stored within the database can subsequently be utilized in later stages of the processing pipeline. From the database, 1000 randomly selected spectrograms were used to train the model, while a separate set of spectrograms was reserved for the validation process. The dataset contains three classes (Lightbridge, OcuSync 20 and OcuSync 40) which are equally represented through the experimental phase. The model was trained to identify characteristic patterns of RF communication signals in the spectrograms

and to distinguish the communication signals associated with different drone types. An additional set of 5000 randomly selected spectrograms was used for inference testing.

The third segment involves applying the statistical analysis introduced in [11] to the existing emissions presented in the output spectrograms obtained in real time from the RF receiver. In this phase, spectrograms that do not contain meaningful information, typically associated with background or ambient noise, are discarded. The ESD threshold is determined empirically, and the threshold defined in [9] was adopted. In addition, an analysis of energy-based signal detection (ESD) was conducted within the Edge AI framework to further assess its role as a pre-filtering mechanism. Different operational scenarios were considered in order to evaluate the ratio between discarded and useful spectrograms, as well as its impact on processing time and overall system performance. The objective was to determine whether such filtering contributes to more efficient resource utilization on the edge platform by reducing unnecessary computations, or whether it introduces limitations that affect its applicability in real-time processing.

The fourth and final segment focuses on the evaluation of the developed YOLO-based solution implemented on the NVIDIA Jetson Orin NX 16GB Edge AI platform. The experiments were performed using the PyTorch-based implementation of the YOLO model. YOLO model was used in the PyTorch (.pt) format through the Ultralytics YOLO library. The model was executed directly within the PyTorch environment, without prior conversion to optimized inference formats such as ONNX or TensorRT.

The Jetson Orin is a powerful and compact AI computing module developed by NVIDIA, designed for Edge AI and robotics applications that offers strong performance for AI inferencing, suitable for real-time applications like computer vision, NLP, and sensor fusion. The device is equipped with 16 GB of 128-bit LPDDR5 memory with a peak bandwidth of up to 102 GB/s, an 8-core Arm Cortex-A78AE CPU and an NVIDIA Ampere GPU featuring 1024 CUDA cores and 32 Tensor Cores, with experiments conducted using Python 3.10.12, PyTorch 2.5, CUDA 12.4 and Ultralytics 8.3.91. The platform supports configurable power modes of 10 W, 15 W, and 25 W, enabling the computational resources to be adapted to the requirements of the deployed application [12].
The objective of this analysis was to assess the feasibility and efficiency of deploying the model in real-time C-UAS applications by evaluating inference performance, neural network output, and hardware resource utilization. This evaluation enables the identification of potential advantages and limitations of edge-based processing compared to traditional server-based or cloud approaches.

The performance of the YOLOv8 model executed through the Ultralytics framework was analyzed using inference and hardware-related metrics. The inference evaluation included model loading time, average processing time per image, processing speed expressed in frames per second (FPS), and total execution time. During the inference, a batch size of 1 was used, meaning that each spectrogram was loaded and processed individually. The confidence threshold was set to 0.5 based on the analysis of the trained model, which indicated that this threshold provides an appropriate balance between precision and recall. At a confidence threshold of 0.5, both precision and recall were approximately 0.995 across all classes, indicating that the model successfully detected the relevant samples while maintaining a minimal number of false-positive and false-negative predictions.

Additionally, hardware performance was assessed through GPU and CPU utilization, power consumption, and GPU temperature to evaluate workload distribution, energy efficiency, and thermal stability. All parameters were recorded during model execution and analyzed across different experimental scenarios, including cases with and without energy-based detection. The obtained results were used to determine whether energy-based filtering contributes to more efficient data processing and improved resource utilization on the Edge AI platform.

4. RESULTS

4.1 Practical evaluation

Practical evaluation was conducted according to the methodology defined in Section 3. The experiments were performed using a dataset of 2000 spectrograms containing ambient signals and 3000 spectrograms with drone communication samples. Two scenarios were analyzed: inference without energy-based filtering and inference with energy-based filtering. Each scenario was evaluated through three test configurations representing different operational conditions. The first configuration considered offline processing, where spectrograms were loaded sequentially from the dataset without additional timing constraints. The second configuration simulated the real-time operation of the RF acquisition system by introducing a 300 ms spectrogram generation interval, corresponding to the actual data flow from the receiver. The third and most demanding configuration involved five simultaneous inference processes, representing a multi-channel operational scenario in which the RF receiver generates spectrograms from five different frequency bands in parallel, which will be case in the future experiments (four different frequency bands are currently in use).

The collected results were stored in tabular form and used for further analysis. Each measurement was repeated ten times, and the presented values represent the corresponding average results, reducing the influence of random variations.

The Table 1 presents the inference results obtained without ESD, based on the parameters defined in the previous section.

Table1. Inference results obtained without ESD

Loading spectrograms scenario	Random	Real-time 300 ms	5 simultaneous processes
Number of spectrograms	5000	5000	5000 x5
Total detection number	5935	5935	5935 x5
Model loading and GPU transfer time	0.298 s	0.201 s	0.235 s
Average time per image	0.046 s	0.042 s	0.111s
Total inference time	225 s	215 s	550 s

The results presented in Table 1 confirm that the implemented YOLO-based classification system achieves real-time performance on the NVIDIA Jetson Orin Edge AI platform. The average inference time per spectrogram in the first test scenario was 46 ms, which enables continuous operation without processing delays. Considering that the current RF acquisition system, described in the Methodology section, generates a new spectrogram every approximately 300 ms, the obtained inference time is significantly shorter than the available processing interval. Therefore, the proposed Edge AI implementation can process incoming data without introducing additional latency.

In the multi-process scenario with five parallel inference streams, the average processing time increased to 111 ms per spectrogram due to higher computational demand. However, this value remains below the spectrogram generation interval, demonstrating that the Jetson Orin platform provides sufficient computational capability for real-time drone detection and classification applications.

The total number of detections differs from the total number of spectrograms because the detection process counts individual objects within each image. The RF communication of interest is generally not continuous within a single spectrogram (Figure 2) and may therefore be represented as two or even three separate objects. Consequently, in the first scenario, 5935 detections were obtained from 3000 useful spectrograms. The total number of detections is expected to be approximately comparable between the two scenarios, with possible differences arising from occasional errors in rejecting relevant spectrograms during the ESD filtering process. In the second scenario, 5920 detections were obtained, since the corresponding to the certain number of rejected spectrograms.

In the second scenario, ESD was introduced as a preprocessing step, while the inference procedure remained unchanged (Table 2). The YOLO inference time remained constant; however, the total processing time increased from 46 ms to 67 ms per spectrogram due to the additional ESD operation, indicating that the energy detection stage requires approximately 20 ms per input.

Although ESD introduces additional computational overhead, its main advantage is the reduction of unnecessary inference operations and more efficient utilization of available hardware resources. In the evaluated dataset, 1965 spectrograms without relevant RF activity were discarded before entering the detection model, thereby reducing unnecessary GPU utilization, memory access, and computational load.

Table 2. Inference results obtained with enabled ESD

Loading spectrograms scenario	Random	Real-time 300 ms	5 simultaneous processes
Number of spectrograms	5000	5000	5000 x5
Number of rejected spectrograms	1965	1965	1965
Total detection number	5920	5920	5920 x5
Model loading and GPU transfer time	0.288 s	0.195 s	0.222 s
Average time per image with ESD	0.067 s	0.065 s	0.105 s
Total inference time	203 s	197 s	319 s

This approach is particularly beneficial in scenarios with a significant amount of background data, where processing every input through the deep learning model would result in increased resource consumption without improving classification performance. For the current acquisition system, which generates spectrograms at approximately 300 ms intervals, the additional ESD processing time remains within acceptable limits. Therefore, energy-based filtering represents an effective optimization strategy for edge-based drone detection systems, enabling more efficient resource utilization while maintaining real-time operation.

4.2 Hardware Performance Analysis

During YOLO model inference on the NVIDIA Jetson Orin platform, real-time monitoring of hardware parameters was performed to evaluate system behavior and resource utilization. The analysis included GPU and CPU utilization, power consumption, operating frequencies, and component temperatures.

The collected measurements were used to assess workload distribution between processing units, energy efficiency, and thermal stability under different experimental scenarios. All parameters were recorded during model execution on spectrogram data, and the results represent averaged values obtained from repeated

measurements.

The hardware performance evaluation, including GPU and CPU utilization, power consumption, memory usage, and temperature variations, is presented in Table 3 for the two most computationally demanding scenarios. The analysis considers five parallel inference processes executed without ESD and with the additional ESD stage, enabling a direct comparison of the system behavior under maximum workload conditions.

Table 3. Hardware performance evaluation

Parameter	Without ESD	With ESD
Average RAM utilization (GB)	7.8	7.8
Maximum RAM utilization (GB)	8.0	7.9
Average GPU utilization (%)	25	30-35
Maximum GPU utilization (%)	98	99
GPU frequency (MHz)	305–407	305–611
Average CPU utilization (%)	28	13
Maximum CPU utilization (%)	89	90
Average power consumption (W)	8	8.4
Maximum power consumption (W)	9.6	9.1
CPU temperature (°C)	55–56	55–56.5
GPU temperature (°C)	51–53	52–53
Swap memory usage (MB)	2	0

Without ESD, the system maintained stable operation during continuous inference with five parallel processes. The average RAM utilization was approximately 7.8 GB, while GPU utilization varied depending on inference activity, reaching peaks close to 98%. The average power consumption remained around 8 W, with short peaks below 10 W. CPU and GPU temperatures remained stable, indicating that the workload was successfully executed within the thermal limits of the Jetson Orin platform.

Compared with the ESD scenario, the main difference is related to the distribution of the computational workload rather than to thermal behavior. The ESD approach reduces unnecessary inference operations by filtering spectrograms with no information before YOLOv8 inference. This reduces the number of redundant neural network executions, while introducing additional preprocessing operations and associated CPU workload. Therefore, the effect of ESD is expected to be reflected primarily in the utilization and distribution of computational resources rather than in a direct reduction of average power consumption.

Additional analysis of system parameters confirms that both approaches are suitable for deployment on the Jetson Orin platform. In the ESD scenario, average CPU utilization decreased from approximately 28% to 13%, while RAM utilization remained practically unchanged at approximately 7.8 GB. In contrast, average GPU utilization increased moderately from approximately 25% to 30–35%, while the maximum GPU utilization remained comparable, reaching approximately 99%. The GPU frequency also reached higher values in the ESD scenario, with a maximum of 611 MHz compared with 407 MHz without ESD. These results indicate that ESD changes the distribution of the computational workload rather than producing a uniform reduction in overall resource utilization. Regarding power consumption, the average value remained at a similar level, changing from approximately 8 W without ESD to 8.4 W with ESD. However, the maximum power consumption decreased from 9.6 W to 9.1 W. The thermal behavior remained stable in both configurations, with CPU and GPU temperatures remaining within comparable ranges. Although ESD introduces additional preprocessing time and computational overhead, it reduces redundant neural network executions and provides an additional mechanism for optimizing the processing pipeline during continuous operation. On the other hand, the scenario without ESD demonstrated that the system is capable of handling continuous inference of five parallel processes, with the average GPU utilization close to 35% and power consumption remaining within the platform limits. Therefore, both configurations satisfy the operational requirements, while the ESD approach provides additional optimization potential for long-term autonomous operation and energy-constrained applications.

5. CONCLUSION

This paper presented the implementation and evaluation of a YOLOv8 deep learning model for drone communication classification using RF spectrogram representations on an Edge AI platform. The proposed approach combines RF signal acquisition, spectrogram representation, deep learning inference, and hardware optimization, forming a complete processing chain for C-UAS applications. The experimental evaluation on the NVIDIA Jetson Orin platform demonstrated the feasibility of executing CNN-based drone detection directly on embedded hardware. The obtained results confirmed that the proposed solution meets the real-time processing requirements, with stable operation even in the most demanding scenario involving five parallel inference processes.

The analysis of hardware parameters showed that the system operates reliably within the constraints of an embedded AI platform, maintaining acceptable resource utilization, power consumption, and thermal behavior.

Furthermore, the integration of ESD as a preprocessing stage reduced the number of unnecessary neural network executions and significantly lowered the average CPU utilization, while maintaining a comparable average power consumption and stable thermal behavior. The results therefore demonstrate that ESD can optimize the processing pipeline without compromising the real-time operation of the system.

Overall, the results confirm that the combination of RF spectrogram analysis, deep learning inference, and Edge AI optimization represents a suitable approach for low latency and resource-efficient drone classification in C-UAS applications. Future work will focus on expanding the RF dataset, improving model generalization, and further optimizing the processing pipeline for more complex operational scenarios.

References

[1] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," *Sensors*, vol. 24, no. 1, p. 125, Dec. 2023, doi: 10.3390/s24010125.

[2] B. Sazdić-Jotić, I. Pokrajac, J. Bajčetić, B. Bondžulić, and D. Obradović, "Single and Multiple Drones Detection and Identification Using RF-Based Deep Learning Algorithm," *Expert Systems with Applications*, vol. 187, p. 115928, Jan. 2022, doi: 10.1016/j.eswa.2021.115928.

[3] M. Al-Sa'D, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open-source drone database," *Future Generation Computer Systems*, vol. 100, pp. 86–97, Nov. 2019.

[4] S. Al-Emadi and F. Al-Senaid, "Drone detection approach based on radio frequency using convolutional neural network," in *Proc. 2020 IEEE Int. Conf. on Informatics, IoT, and Enabling Technologies (ICIoT)*, Doha, Qatar, Feb. 2020, pp. 29–34.

[5] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and Bluetooth interference," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 60–76, 2020, doi: 10.1109/OJCOMS.2019.2955889.

[6] A. Garcia-Perez, R. Miñón, A. I. Torre-Bastida, and E. Zulueta-Guerrero, "Analysing edge computing devices for the deployment of embedded AI," *Sensors*, vol. 23, no. 23, p. 9495, 2023, doi: 10.3390/s23239495.

[7] R. Singh, S. S. Gill, "Edge AI: A Survey", *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71–92, 2023, doi: 10.1016/j.iotcps.2023.04.001.

[8] N. W. Khan, M. S. Alshehri, M. A. Khan, S. Almakdi, N. Moradpoor, A. Alazeb, S. Ullah, N. Naz, and J. Ahmad, "A hybrid deep learning-based intrusion detection system for IoT networks," *Mathematical Biosciences and Engineering*, vol. 20, no. 8, pp. 13491–13520, 2023, doi: 10.3934/mbe.2023602.

[9] V. Joksimović, "Design and Implementation of a Convolutional Neural Network for Drone Classification Based on Radio Communication", Master's Thesis, University of Defence in Belgrade, Military Academy, Belgrade, 2026.

[10] T. Šević, V. Joksimović, I. Pokrajac, B. Radiana, B. Sazdić-Jotić, D. Obradović, "Interception and detection of drones using RF-based dataset of drones", Scientific Technical Review, vol. 70, no. 2, pp. 29–34, 2020, doi: 10.5937/str2002029S.

[11] S. Zurovac, N. Petrović, V. Joksimović, I. Pokrajac, D. Mikanović, B. Sazdić-Jotić, "Simple energy detector for two-stage classification for antidrone systems", in 11th International Scientific Conference on Defensive Technologies – OTEH, Serbia, 09-11 Oct. 2024, pp. 369–375, doi: 10.5937/OTEH24066Z.

[12] H. V. Pham, T. G. Tran, C. D. Le, A. D. Le, H. B. Vo, "Benchmarking Jetson Edge Devices with an End-to-end Video-Based Anomaly Detection System", in Proc. FICC 2024, Berlin, Germany, 4–5 Apr. 2024, pp. 358–374, doi: 10.1007/978-3-031-53963-3_25.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

METHODS FOR IMPROVING MACHINE LEARNING MODELS FOR PREDICTING MULTI-HOLE PRESSURE PROBE CHARACTERISTICS

IVANA PETKOVIĆ
Military Technical Institute, Belgrade, ivana.petkovic.vti@gmail.com

SNEŽANA ZUROVAC
Military Technical Institute, Belgrade, zurovac@medianis.net

DIJANA DAMLJANOVIĆ
Military Technical Institute, Belgrade, didamlj@gmail.com

ILIJA NENADIĆ
Military Technical Institute, Belgrade, ilijanenadic98@gmail.com

Abstract: *The standard engineering process for developing a multi-hole pressure probe included the use of 3D software during the design phase. At the same time, CFD (Computational Fluid Dynamics) was used to calculate the main characteristics of the probe. After the probe geometry was selected, the CFD results were verified by wind tunnel testing. This process ensured that the predicted probe characteristics were accurate under real flow conditions.*

However, numerical simulations and experimental testing required a significant amount of time and financial resources. For this reason, a machine learning model for predicting the characteristics of a multi-hole pressure probe was developed at the Military Technical Institute in Belgrade [1]. The selected model was a Random Forest Regressor because it achieved the lowest MAE (Mean Absolute Error).

The model was trained to predict probe characteristics for subsonic, transonic, and supersonic flow regimes and for two probe geometries: a sharp probe and a blunt probe. In the previous study, the Random Forest model achieved an MAE of 0,001681.

The aim of this research was to improve the performance of the existing model by using machine learning optimization methods. In addition, the effect of the dataset on model performance was analyzed. Grid Search with cross-validation was applied to find the best hyperparameter combination and improve the prediction accuracy of the model for different flow regimes.

Keywords: *multi-hole pressure probe, machine learning, Random Forest Regressor, hyperparameter optimization, flow regimes.*

1. INTRODUCTION

In previous research at the Military Technical Institute in Belgrade, a machine learning model based on the Random Forest Regressor was developed to predict the characteristics of a multi-hole pressure probe. The model was trained using experimental data from wind tunnel tests. The input data included flow parameters and probe geometry characteristics, while the output was ΔP/P0 (the normalized pressure difference) [1].

The previous model achieved a MAE of 0,001681. The Random Forest algorithm was selected as the best model after it was compared with other machine learning models [1]. Although the model showed high prediction accuracy, its performance could still be improved.

Since the performance of a machine learning model depended on the selected hyperparameters, this paper investigated different methods for improving the existing Random Forest model. The main focus was on hyperparameter optimization using Grid Search together with cross-validation to find the best combination of hyperparameters that gave the lowest prediction error.

In addition, the effect of dataset segmentation and dataset characteristics on model performance was analyzed. The analysis included different flow regimes (subsonic, transonic, and supersonic) and two different probe geometries. The results showed that model accuracy depended not only on the number of samples but also on the data distribution and homogeneity within each flow regime.

The aim of this paper was to improve the existing Random Forest model through hyperparameter optimization and analysis of the dataset structure.

Corresponding author: I. Petković;
E-mail: ivana.petkovic.vti@gmail.com:
DOI: 10.5937/OTEH26063P

2. DATASET DESCRIPTION

The dataset used in this study contained a total of 2095 samples collected for two probe geometries under different flow regimes. The samples were collected during wind tunnel tests in the T-35 and T-38 wind tunnels at the Military Technical Institute. The sharp probe is shown in Figure 1, while the blunt probe is shown in Figure 2. Each sample represented one measurement and included the following input parameters: static pressure (Pst), Mach number (M), flow temperature (T0), and effective angle of attack (θ_eff). The output variable was the normalized pressure difference (ΔP/P0).

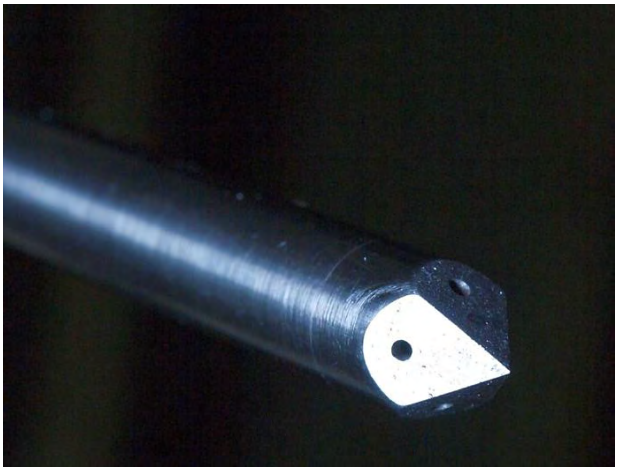


Figure 1. Four-hole pyramidal pressure probe, declared as *a sharp probe*; tested in the VTI T-38 wind tunnel in the Mach number range from 0.2 to 4

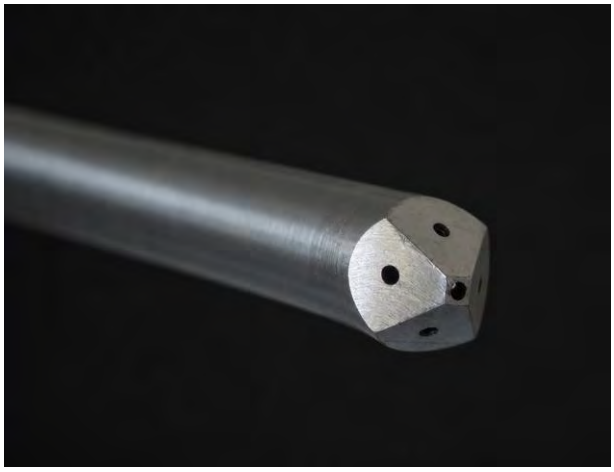


Figure 2. Five-hole pyramidal pressure probe designed for small UAV (unmanned aerial vehicle), declared as *a blunt probe*; single tested in the VTI T-38 wind tunnel

Data for the sharp probe were available for all three flow regimes (subsonic, transonic, and supersonic), while data for the blunt probe were available only for the subsonic flow regime. To better analyze the effect of the flow regime and probe geometry on model performance, the complete dataset was divided into four subsets. The number of samples in each subset is shown in Table 1.

The largest subset contained 1025 samples for the sharp probe in the supersonic flow regime, while the smallest subset contained 84 samples for the sharp probe in the transonic flow regime.

Table 1. Dataset structure

Dataset	Flow Regime	Geometry	Samples
sharp_subsonic	subsonic	sharp	672
sharp_transonic	transonic	sharp	84
sharp_supersonic	supersonic	sharp	1025
blunt_subsonic	subsonic	blunt	314

3. METHODOLOGY

3.1 Random Forest Regressor

In the previous study, the Random Forest Regressor was selected as the best model for predicting the characteristics of the pressure probe because it achieved the lowest MAE [1]. Based on these results, the same model was used in this study as the starting point for further optimization.

Random Forest is an ensemble machine learning method that consists of many decision trees. During training, each tree is built using a different subset of the training data. The final prediction is obtained by averaging the predictions of all trees. This approach reduces overfitting and improves the model's ability to make accurate predictions on new data [2].

To explain how the Random Forest algorithm works, Figure 3 shows one decision tree generated using the complete training dataset from this study. The maximum depth of the tree is limited to 3.

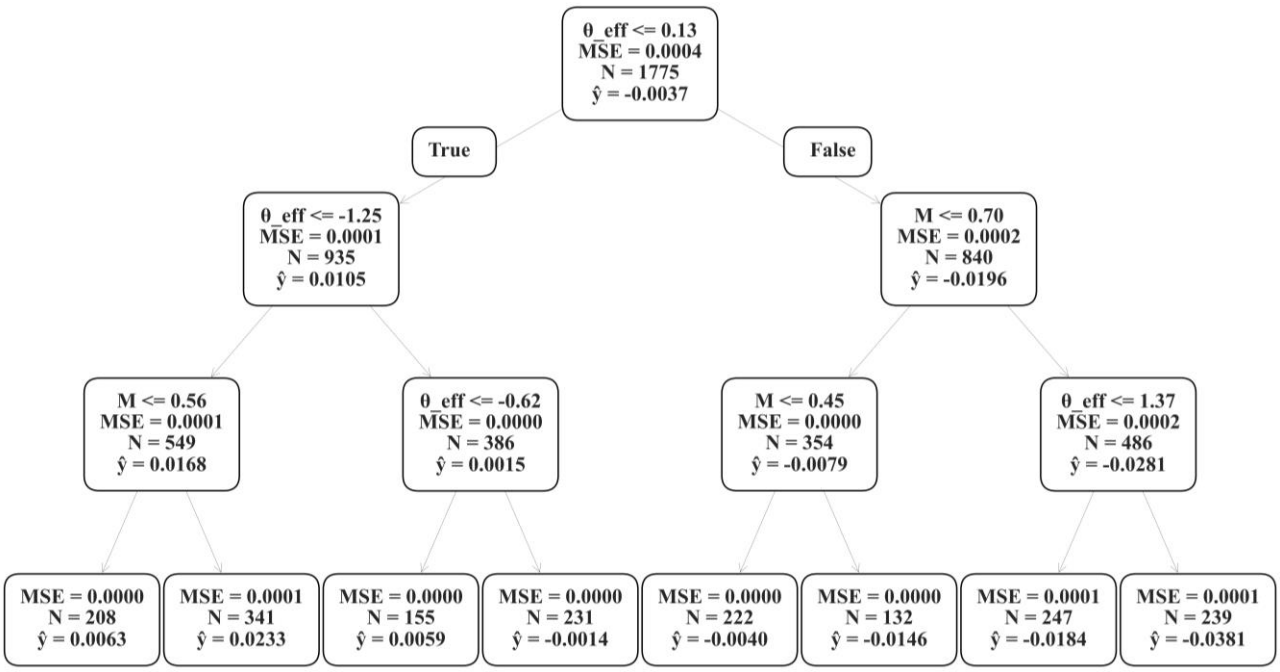


Figure 3. Example of a decision tree generated for the complete dataset

The root node represents the first data split. In this example, the algorithm selected the effective angle of attack, θ_eff, as the first splitting criterion. The threshold θ_eff ≤ 0.13° was automatically selected during training because it produced the largest reduction in the Mean Squared Error (MSE) among all possible splits of all input features. To find the best split, the algorithm checks different threshold values from the training data for each input feature. For every possible split, it calculates the MSE of the two new groups. The feature and threshold that give the lowest total MSE are selected.

All training samples are divided into two groups depending on whether θ_eff ≤ 0.13° or θ_eff > 0.13°. The root node contains all training samples used to train the model.

After the first split, the algorithm repeats the same process for each new group of data. At every node, it checks all input features and different threshold values again. The split that gives the largest reduction in MSE is selected. In this example, the next splitting criteria are again θ_eff and the Mach number (M), but with different threshold values. For example, on the right branch the next split is M ≤ 0.70, while on the left branch θ_eff is selected again with the threshold −1.25°. This shows that the algorithm chooses the best feature and threshold separately at each node.

Each node contains the Mean Squared Error (MSE), the number of training samples (N), and the predicted mean value of the target variable (ŷ). The parameter N represents the number of training samples contained in the node, while ŷ represents the average value of the target variable ΔP/P0 for those samples. The MSE shows how similar the target values are within a node. Lower MSE values indicate that the target values are more similar, meaning that the node is more homogeneous.

The splitting process ends at the leaf nodes. Each leaf contains training samples with similar input values. The prediction of a single decision tree is the value **ŷ**, which represents the average value of the target variable ΔP/P0 for all training samples contained in that leaf. The final prediction of the Random Forest Regressor is obtained by averaging the predictions of all decision trees in the forest.

3.2 Hyperparameter optimization

To improve the performance of the existing Random Forest model, hyperparameter optimization was performed using the Grid Search method. Grid Search tested all predefined combinations of hyperparameters to find the configuration that gave the lowest MAE [3].

In this study, the effect of three main hyperparameters of the Random Forest algorithm was analyzed: the number of trees in the forest (n_estimators), the maximum tree depth (max_depth), and the minimum number of samples required to split a node (min_samples_split). The tested hyperparameter values are shown in Table 2.

The selected hyperparameter values and different cross-validation configurations were combined in ten experiments to find the best model configuration. The cross-validation procedure used in this study is described in the next section.

Table 2. Tested hyperparameters

Parameter	Values
n_estimators	200, 300, 500, 700, 800, 1000
max_depth	10, 20, 30, 40, None
min_samples_split	2, 5, 7, 9, 10

3.3 Cross-validation

The performance of the model was evaluated using k-fold cross-validation. In this method, the dataset is divided into k subsets of approximately equal size. During the validation process, the model is trained using k−1 subsets, while the remaining subset is used for testing. This process is repeated k times, so that each subset is used once as the validation set [4].

In this study, cross-validation was used as part of the Grid Search procedure to select the best combination of hyperparameters for the Random Forest Regressor. In addition to hyperparameter optimization, the effect of the number of validation folds on model performance was also analyzed. For this purpose, ten experiments were performed using different values of the cv parameter (5, 10, 15, 40, 50, 60 and 75).

For each combination of hyperparameters, the average MAE was calculated over all validation iterations. The configuration with the lowest MAE was selected as the best model configuration.

3.4 Dataset segmentation

To study the effect of the dataset on the performance of the Random Forest regressor, the complete dataset was divided into four separate subsets. The division was based on the probe geometry and the flow regime. The four subsets were sharp subsonic, sharp transonic, sharp supersonic, and blunt subsonic.

The same Random Forest model was used for all four subsets. The model used the optimal hyperparameters found during the optimization process. The model was evaluated using k-fold cross-validation. This made it possible to compare the results under the same conditions.

The performance of each model was measured using the MAE. The Mach number distribution was analyzed for the sharp supersonic and blunt subsonic datasets because they had the highest and the lowest MAE. The distribution of the target variable ΔP/P0 was analyzed for all four dataset subsets using boxplot diagrams. The range and the standard deviation (Std) of the target variable were also calculated to measure the data variability.

The goal of this analysis was to find out whether the prediction accuracy depended only on the number of samples or also on the data distribution and variability.

4. RESULTS

4.1 Hyperparameter optimization results

The results of the hyperparameter optimization are shown in Table 3. Compared with the previously developed model [1], the optimized Random Forest Regressor achieved a lower MAE. The best result was obtained with the following hyperparameter values: n_estimators = 500, max_depth = 10, min_samples_split = 2, and cv = 50.

The obtained MAE was 0,001601, which represented an improvement of about 4.8% compared to the previous study, where the MAE was 0,001681. The results showed that hyperparameter optimization improved the prediction accuracy of the model, although the improvement was relatively small. This indicated that the original model already had high accuracy, while the optimization further improved its performance. All ten experiments were performed on the complete dataset.

Table 3. Comparison of MAE before and after hyperparameter optimization

Model	MAE
Random Forest (previous work)	0,001681
Random Forest (Grid Search optimized)	0,001601

4.2 Dataset segmentation results

The prediction results for the four dataset subsets are shown in Table 4. Each subset was analyzed separately using the Random Forest model with the same optimal hyperparameters (n_estimators = 500, max_depth = 10, and min_samples_split = 2). The model was evaluated using 50-fold cross-validation.

Table 4. Model performance for different dataset subsets

Dataset	Samples	MAE
sharp_subsonic	672	0,000804
sharp_transonic	84	0,001565
sharp_supersonic	1025	0,002799
blunt_subsonic	314	0,000405

The lowest MAE was obtained for the blunt subsonic dataset and was 0,000405. The highest MAE was obtained for the sharp supersonic dataset and was 0,002799. Although the sharp supersonic dataset contained the largest number of samples, it produced the highest prediction error. On the other hand, the sharp transonic dataset contained only 84 samples, but its MAE (0,001565) was much lower than the MAE of the sharp supersonic dataset. These results showed that a larger number of samples did not always lead to better prediction accuracy.

The distribution of the Mach number for the datasets with the highest and the lowest MAE is shown in Figure 4. It can be seen that the sharp supersonic dataset covered a much wider Mach number range than the blunt subsonic dataset. This indicated a larger variation of flow conditions in the sharp supersonic dataset.

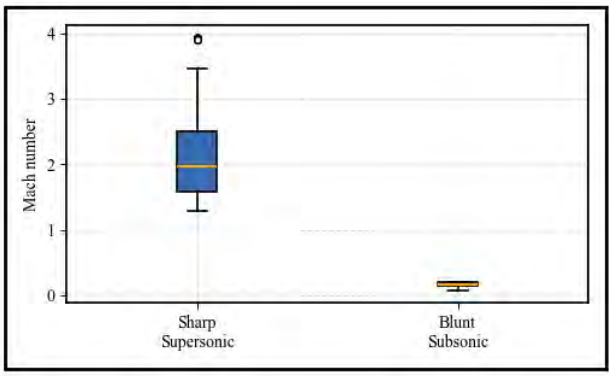


Figure 4. Distribution of Mach Number

Figure 5 shows the distribution of the target variable ΔP/P0 for all four dataset subsets. The differences between the distributions showed that the datasets had different levels of data variability.

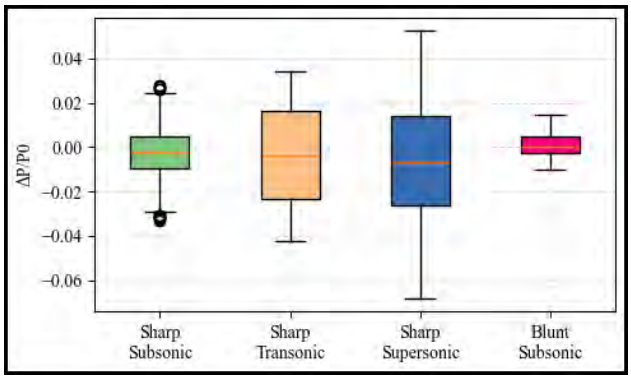


Figure 5. Distribution of ΔP/P0

For a more detailed analysis, the range, Std and MAE were calculated for each dataset subset. The results are shown in Table 5.

Table 5. Statistical characteristics of the dataset subsets

Dataset	Samples	Range	Std	MAE
sharp_subsonic	672	0,0614	0,0119	0,000804
sharp_transonic	84	0,0764	0,0225	0,001565
sharp_supersonic	1025	0,1203	0,0259	0,002799
blunt_subsonic	314	0,0249	0,0052	0,000405

The sharp supersonic dataset had the largest range (0,1203), the highest standard deviation (0,0259), and the highest MAE (0,002799). In contrast, the blunt subsonic dataset had the smallest range (0,0249), the lowest standard deviation (0,0052), and the lowest MAE (0,000405).

The results showed that the prediction accuracy of the Random Forest model did not depend only on the number of samples. Higher data variability and a wider data distribution made the learning process more difficult and increased the prediction error. On the other hand, more homogeneous datasets allowed the model to make more accurate predictions and achieve lower MAE values.

5. CONCLUSION

This paper analyzed different methods for improving the Random Forest Regressor developed for predicting the characteristics of a multi-hole pressure probe. By using Grid Search optimization and different cross-validation configurations, the MAE was reduced by about 4.8%. The best results were achieved with the following hyperparameter values: n_estimators = 500, max_depth = 10, min_samples_split = 2, and cv = 50.

In addition to hyperparameter optimization, the effect of dataset characteristics on model performance was also analyzed. The results showed that the sharp supersonic dataset, although it contained the largest number of samples, had the highest MAE and the highest variability of the target variable. In contrast, the blunt subsonic dataset had the lowest variability and the lowest MAE.

The results showed that, besides the number of samples, the data distribution and variability also had a significant effect on the accuracy of the Random Forest Regressor. They also showed that the structure of the dataset was an important factor that affected model performance. Therefore, future research will focus on a more detailed analysis of dataset characteristics and organization, as well as on developing methods for creating more representative training datasets.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] Zurovac S., Damljanović D., Nenadić I., Vuković Đ.: *Development of a Predictive Model for Aerodynamic Probe Design Based on Wind Tunnel Test Data*, 60th 3AF International Conference on Applied Aerodynamics, Paris - France, 2026.

[2] Breiman L.: *Random Forests*, Machine Learning, Vol. 45, No. 1, pp. 5–32, 2001.

[3] Pedregosa F., Varoquaux G., Gramfort A., et al.: *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, Vol. 12, pp. 2825–2830, 2011.

[4] Kohavi R.: *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*, Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1137–1143, 1995.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PERFORMANCE EVALUATION OF PSO TUNED FRACTIONAL ORDER ILC FOR THE LOLA15 ROBOT

NIKOLA ŽIVKOVIĆ
Lola Institute Ltd., Kneza Višeslava 70a 11030 Belgrade, nikola.zivkovic@li.rs

MIHAILO LAZAREVIĆ
Faculty of Mechanical Engineering, Kraljice Marije 16 11060 Belgrade, mlazarevic@mas.bg.ac.rs

JELENA VIDAKOVIĆ
Lola Institute Ltd., Kneza Višeslava 70a 11030 Belgrade, jelena.vidakovic@li.rs

Abstract: *Due to uncertainty in their mathematical models and external disturbances, which affect the quality of model-based control, the trajectory-tracking performance of serial robotic manipulators is always in question. There are many methods for compensating for undesired uncertainties and external disturbances, but for a system that operates repeatedly, such as a robot manipulator in an industrial setting, the Iterative Learning Control (ILC) strategy is particularly effective. Additional improvement is achieved by integrating fractional calculus into ILC, leading to a Fractional-order Iterative Learning Controller (FOILC). The FOILC controller iteratively compensates for the influence of parametric uncertainty on trajectory tracking. Optimal performance is achieved by tuning control parameters using the Particle Swarm Optimization (PSO) algorithm, which compares multiple objective functions to determine the most effective performance criterion. The efficiency of the proposed approach is validated through simulation on the RL15 6-DOF industrial serial robot. The results demonstrate the performance superiority of the tuned FOILC over its integer-order counterpart, while the comparative analysis of tuning criteria identifies the optimal objective function for achieving the best tracking accuracy.*

Keywords: *iterative learning control, fractional calculus, optimization, performance criteria, industrial robots.*

1. INTRODUCTION

Articulated robots have an open-loop kinematic chain structure, for which trajectory tracking is a fundamental requirement in applications such as welding, assembly, and sorting, and it poses a challenge for control design, given the system's coupled, nonlinear nature, unmodeled dynamics, external disturbances, and other uncertainties.

Today, PID controllers remain a primary feedback solution for most control problems, including robot manipulators [1], and a PD position controller combined with a nonlinear control strategy is particularly well-suited for articulated robots [2]. Rather than relying on a model-based feedback law such as inverse dynamics control, an Intelligent PID controller is adopted here as the feedback controller, preserving the model-free aspect of the overall control scheme while still providing a satisfactory response [3]. Articulated robot manipulators are generally used for repetitive tasks such as welding, assembly, or sorting. Iterative Learning Control (ILC), a memory-based learning control strategy [4], excels at improving the transient response of systems operating repetitively over a finite time interval [5], without relying on precise knowledge of the process model [6]. ILC can be easily incorporated into existing control systems [7].

Incorporation of fractional calculus theory (succinctly put, a generalization of integer-order calculus [8]) into ILC results in Fractional Order Iterative Learning Control (FOILC) [9]. The main advantage of fractional-order controllers is that they offer more degrees of freedom than their integer-order counterparts, thereby yielding better performance [10]. An increased number of control system parameters requires tuning for better system response, and the choice of performance criterion used in that tuning process, such as the Integral of Squared Error (ISE), Integral of Time-weighted Squared Error (ITSE), Integral of Absolute Error (IAE), or Integral of Time-weighted Absolute Error (ITAE), directly affects the resulting transient response. Although fractional-order PID tuning has been widely studied in the literature (e.g., [11], [12]), a proper tuning procedure that accounts for these performance criteria is still lacking for FOILC controllers. The research in [13] addresses the problem of tuning FOILC control parameters for a linearized robot with no uncertainty. Another research tackled tuning FOILC parameters for an articulated robot with model parameter uncertainty [14]. The main motivation for this research is to find an easy, computationally efficient, and performance-improving solution to tuning FOILC parameters by comparing the resulting closed-loop performance across several standard error-based criteria. The Particle Swarm Optimization (PSO) was chosen for FOILC parameter tuning, with the objective function formulated around the aforementioned performance criteria.

Corresponding author: N. Živković;
E-mail: nikola.zivkovic@li.rs

DOI: 10.5937/OTEH26064Z

This paper focuses on applying FOILC to the legacy industrial robot LOLA15 (RL15) with an Intelligent PID position controller. This paper consists of the following sections: Section II provides a brief introduction to ILC and describes the controller design, divided into subsections on the Intelligent PID, the Fractional Order Iterative Learning Controller, and the tuning of FOILC control parameters using PSO with respect to different performance criteria (ISE, ITSE, IAE, ITAE); Section III presents the simulation results on the RL15 robot; and Section IV presents the concluding remarks.

2. CONTROL SYSTEM DESIGN

2.1. Intelligent PID

The intelligent PID (i-PID) control strategy is based on the combination of classical PID and dynamics estimation via ultra-local model [3]. The general form of the i-PID control law for SISO (Single Input Single Output) systems is:

$$u(t) = b^{-1}\left(r^{(v)}(t) - \hat{f}(t) + PID(e)\right), \qquad (1)$$

where $r^{(v)}(t)$ is v-th derivative of the desired output, $\hat{f}(t)$ is an estimate of nonlinear dynamics, model parameter uncertainties and external disturbances, PID(e) is general form of any PID controller structure (P, PI, PD, etc.), and b^{-1} is inverse of control parameter b. The detailed algorithm of obtaining the value of estimate $\hat{f}(t)$ can be found in [3].

Expanding from SISO to MIMO (Multiple Input Multiple Output) n-degree of freedom articulated robotic systems, the control law for i-PID can be written in the following form:

$$\mathbf{u}(t) = \mathbf{B}^{-1}\left(\mathbf{q}_d^{(v)}(t) - \hat{\mathbf{f}}(t) + \mathbf{PID}(\mathbf{e})\right), \qquad (2)$$

where $\mathbf{B}^{-1}$ is inverse matrix of $n \times n$ matrix $\mathbf{B}$, $\mathbf{q}_d^{(v)}(t)$ is a $n \times 1$ vector of v-th derivative of the desired output, $\hat{\mathbf{f}}(t)$ is a $n \times 1$ vector of an estimate of nonlinear dynamics, model parameter uncertainties and external disturbances, and $\mathbf{PID}(\mathbf{e})$ is a general form of any PID controller structure. For articulated robot matrix $\mathbf{B}$ can be obtained based on a robot's inertia matrix as explained in [15]. For this study proposed PID controller structure is adopted as PD and final form of the total control law is given as:

$$\mathbf{u}(t) = \mathbf{B}^{-1}\left(\mathbf{q}_\mathbf{d}^{(v)}(t) - \hat{\mathbf{f}}(t) + \mathbf{p}(t) + \mathbf{v}(t)\right), \qquad (3)$$

where $\mathbf{p}(t)$ is PD controller:

$$\mathbf{p}(t) = \mathbf{K_P}\mathbf{e}(t) + \mathbf{K_D}\dot{\mathbf{e}}(t), \qquad (4)$$

where, $\mathbf{e}_i = \mathbf{q_d} - \mathrm{q}_i$ is $n \times 1$ vector of position errors, $\dot{\mathbf{e}}_i = \dot{\mathbf{q}}_\mathbf{d} - \dot{\mathbf{q}}_i$ is the $n \times 1$ vector of velocity errors, $\mathbf{K_P}$, $\mathbf{K_D}$ are PD gains. Vector $\mathbf{v}(t)$ from (3) is the FOILC controller control law which will be explained in the following subsection.

2.2. Fractional order D^α-type ILC

Here, the FOILC D^α-type controller is considered to improve trajectory tracking and ensure system robustness to uncertainties compared to integer order. The FOILC control input vector $\mathbf{v}(t)$ is defined by the following equation:

$$\mathbf{v}_i = \mathbf{Q}\left(\mathbf{v}_{i-1} + \mathbf{L_P}\,\mathbf{p}_{i-1} + \mathbf{L_F}\,D_{GL}^{\alpha_j}(\mathbf{e}_{i-1})\right), \qquad (5)$$

where i is the iteration index, j is the joint index, $\mathbf{p}_{i-1}(t)$ is the vector of PD control signal, $\mathbf{L_P}$ and $\mathbf{L_F}$ are positive diagonal learning gain matrices, $D_{GL}^{\alpha_j}$ is the fractional derivative operator of the order α_j according to the Griunwald-Letnikov definition, and $\mathbf{Q}$ is positive diagonal robustness matrix. The overview of the proposed controller structure is displayed in Figure 1.

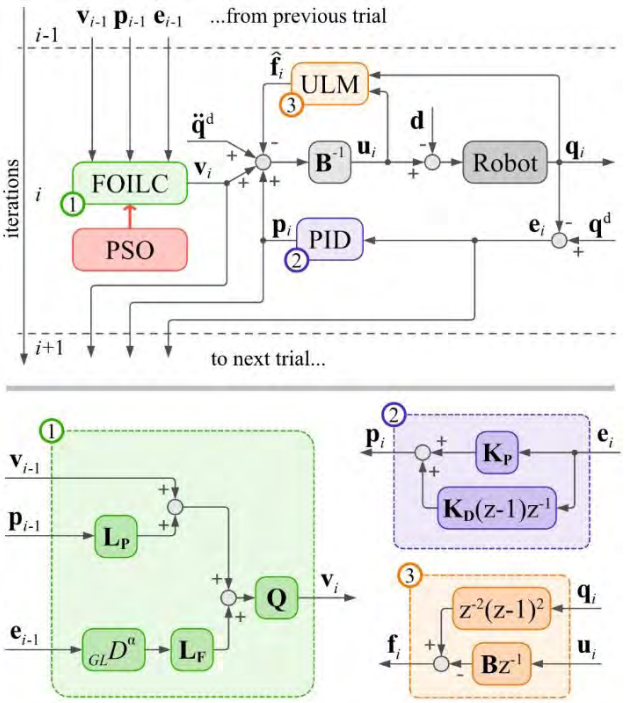


Figure 1. Control scheme of the proposed i-PD FOILC controller for robot LOLA15.

2.3. Tuning procedure using Particle Swarm Optimization

Introducing fractional derivative and integral operators into the control law increases the number of tunable degrees of freedom, which in turn leads to improved performance. Reference [16] demonstrates that raising the fractional order α of the derivative operator accelerates error convergence. This observation motivates treating α as a tunable parameter of the FOILC law. The same reasoning applies to the learning gain matrices $\mathbf{L_P}$ and $\mathbf{L_F}$ that appear in the FOILC control law (5), since both directly shape the system's output and are likewise subject to tuning. In contrast, the elements of the robustness filter matrix $\mathbf{Q}$ is held constant and excluded from the tuning procedure, leaving α, $\mathbf{L_P}$ and $\mathbf{L_F}$ as the only parameters to be optimized.

Unlike classical PID controllers, PID-type ILC laws such as (5) lack a widely accepted tuning methodology. For this reason, the Particle Swarm Optimizer (PSO) algorithm was selected as a straightforward approach to solving the FOILC parameter-tuning problem. PSO is a population-

based, nature-inspired optimization technique modeled on the collective behavior of social organisms such as flocking birds or schooling fish. Since its introduction in [17], the algorithm has become a widely used heuristic for tackling complex optimization tasks, and numerous variants and refinements have emerged over the years. This work adopts the PSO variant described in [18]; the corresponding algorithm parameters are listed in Table 1, most of which were determined empirically. Rather than re-tuning α, $\mathbf{L_P}$ and $\mathbf{L_F}$ after every single iteration as in [13,14], a fixed-horizon tuning approach is adopted: PSO is run once to find a single set of FOILC parameters, with the objective function evaluated over a fixed horizon of i_{max} ILC iterations rather than a single one. Per-iteration tuning was avoided because allowing the gains to vary from one iteration to the next undermines the contraction argument underlying ILC's convergence guarantees and reduces the search to a one-step-ahead optimization, which can drive the error toward divergence rather than zero.

To quantify tracking performance for tuning, standard error-based performance criteria like the Integral of Squared Error (ISE), the Integral of Time-weighted Squared Error (ITSE), the Integral of Absolute Error (IAE), and the Integral of Time-weighted Absolute Error (ITAE) are compared to evaluate which yields the best performance improvement for the proposed control system. The general form of the objective function minimized by PSO, evaluated over the fixed tuning horizon, is given as:

$$C = \sum_{k=1}^{i_{max}} J_k + P_b + P_{sat} + P_m, \tag{6}$$

where i_{max} denotes the duration of the tuning horizon, J_k represents the selected performance-criterion weighting corresponding respectively to ISE, ITSE, IAE, and ITAE:

$$ISE: J_k = \sum_{j=1}^{n} \int_0^T e_j^2(t)dt, \tag{7}$$

$$IAE: J_k = \sum_{j=1}^{n} \int_0^T |e_j(t)|dt, \tag{8}$$

$$ITSE: J_k = \sum_{j=1}^{n} \int_0^T te_j^2(t)dt, \tag{9}$$

$$ITAE: J_k = \sum_{j=1}^{n} \int_0^T t|e_j(t)|dt, \tag{10}$$

Term P_b from (6) is a bound violation penalty applied according to the following cases:

$$P_b = \begin{cases} 0, & L_{P_{jj}}, L_{F_{jj}}, \alpha_j > 0 \\ 1000, & \text{otherwise} \end{cases}. \tag{11}$$

Beyond penalizing particles for violations of tuned parameter bounds, the tuning procedure also enforces the manipulator's joint torque saturation limits as an explicit constraint:

$$P_{sat} = \begin{cases} 0, & |u_j| < [2130, 4500, 2118] \\ 1000, & \text{otherwise} \end{cases}. \tag{12}$$

Through this mechanism, P_{sat} effectively removes from the search process any particles that either destabilize the system or push its joint torques beyond their allowable limits by imposing an additional cost on them. Finally, penalty term P_m penalizes monotonicity violation across ILC iterations:

$$P_m = \begin{cases} 0, & \|\mathbf{e}_{k-1}\| < \|\mathbf{e}_k\| \\ 2 \cdot (\|\mathbf{e}_{k-1}\| - \|\mathbf{e}_k\|), & \text{otherwise} \end{cases}. \tag{13}$$

Table 1. Parameters for the PSO algorithm.

Variable	Value
Population number	30
Iteration number	10
Initial value of the individual-best acceleration factor	0.5
Final value of the individual-best acceleration factor	2.5
Initial value of the global-best acceleration factor	2.5
Final value of the global-best acceleration factor	0.5
Initial value of the inertia factor	0.9
Final value of the inertia factor	0.6

3. SIMULATION RESULTS

The proposed controller is tested through simulation using the LOLA15 industrial robot dynamics model with only the first three degrees of freedom considered. The masses of links 5 and 6 are added to the mass of link 4, since only a 3-DoF structure is considered. The experimental results are out of scope for this study since an experimental study requires additional torque safeguards to be implemented on the real controller. The LOLA15 robot's kinematics and dynamics parameters are listed in [19] and simulation is carried out in Simulink using the Robotics Systems Toolbox and Runge-Kutta solver (ode4) for ordinary differential equations with fixed-step T_s = 0.001s. The desired joint-space trajectories are obtained via inverse kinematics, converting end-effector desired path defined in the Cartesian space as a continuous circular path as illustrated in Figure 2.

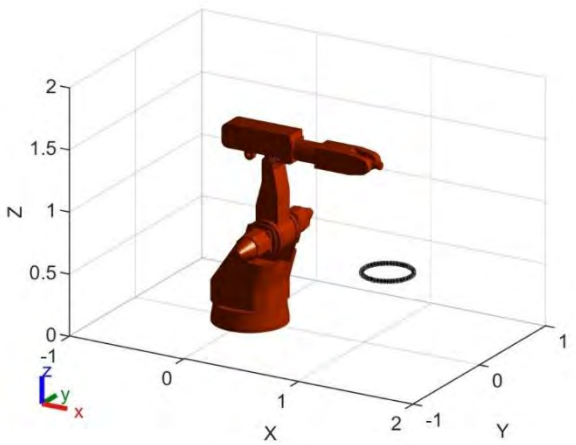


Figure 2. LOLA15 robot and desired path of the end-effector.

The gain matrices $\mathbf{K_P}$ and $\mathbf{K_D}$ of the PD controller are tuned empirically to a non-optimal solution in order to emphasize

the influence of FOILC on potentially under-tuned real systems. FOILC control parameters are tuned as described in subsection 2.3 of this paper and robustness matrix **Q** is set as diag([0.9 0.9 0.9]). The tuned control system parameters are given in Tables 2., 3., 4. and 5. This study measures the ILC controllers' performance quality in the iteration domain with the infinite norm of the error:

$$\|e_j\|_\infty = \max_k |e_j(t)|, \tag{14}$$

Table 2. Tuned parameters of the FOILC controller for ISE criterion.

Parameter	Value
$\mathbf{L_P}$	diag([0.07, 0.41, 0.93])
$\mathbf{L_F}$	diag([0.52, 0.62, 0.55])
α	[0.99, 0.95, 0.47]

Table 3. Tuned parameters of the FOILC controller for IAE criterion.

Parameter	Value
$\mathbf{L_P}$	diag([0.08, 1.17, 0.35])
$\mathbf{L_F}$	diag([0.92, 1.08, 0.52])
α	[0.18, 0.78, 1.64]

Table 4. Tuned parameters of the FOILC controller for ITSE criterion.

Parameter	Value
$\mathbf{L_P}$	diag([0.07, 0.35, 0.34])
$\mathbf{L_F}$	diag([1.15, 0.55, 0.7])
α	[0.89, 1.25, 1.44]

Table 5. Tuned parameters of the FOILC controller for ITAE criterion.

Parameter	Value
$\mathbf{L_P}$	diag([0.09, 1.05, 1.32])
$\mathbf{L_F}$	diag([0.16, 1.33, 0.42])
α	[0.97, 1.89, 1.62]

The measure in (15), graphed for each iteration, shows how much the maximum absolute error of each joint is reduced (or increased if the learning process is unstable) after each iteration.

Based on the adopted quality measure in (15), Figure 3 shows that the FOILC controller improves the system's tracking performance, and for every joint and every criterion, the learning process is stable and convergent, meaning the maximum absolute error decreases over iterations for this setup.

Examining Figure 3 further, it can be seen that for the first joint, there is a minimal difference in performance among the selected criteria. On the other hand, there is a clear distinction between all selected criteria for the maximum absolute error for joint 2. ITSE criterion error reduction is the smallest relative to the other selected criteria, followed by ISE and IAE, and finally, the ITAE criterion shows best performance with the lowest converged maximum absolute error at the 10th iteration.

Maximum absolute error convergence for the joint 3 features situation, where performance is split into two groups: ITAE and ISE have similar error values compared to the IAE and ITSE group, which are almost identical but slower in convergence compared to the former.

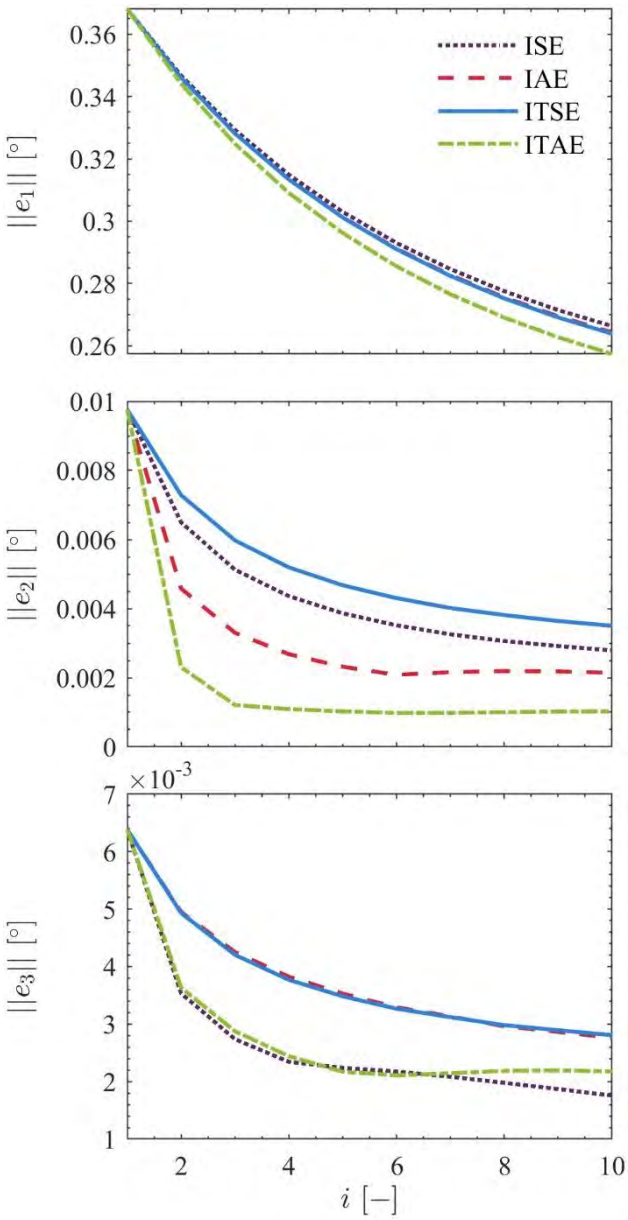


Figure 3. Infinite norm of the joint position error per joint comparison between ISE, IAE, ITSE and ITAE performance criteria.

Figure 4 shows joint position errors and indicate that FOILC enhances performance across iterations and illustrates differences between selected performance criteria in time domain. Figure 4 reflects results shown in Figure 3 in time domain at 10th iteration i.e. at the end of the learning process.

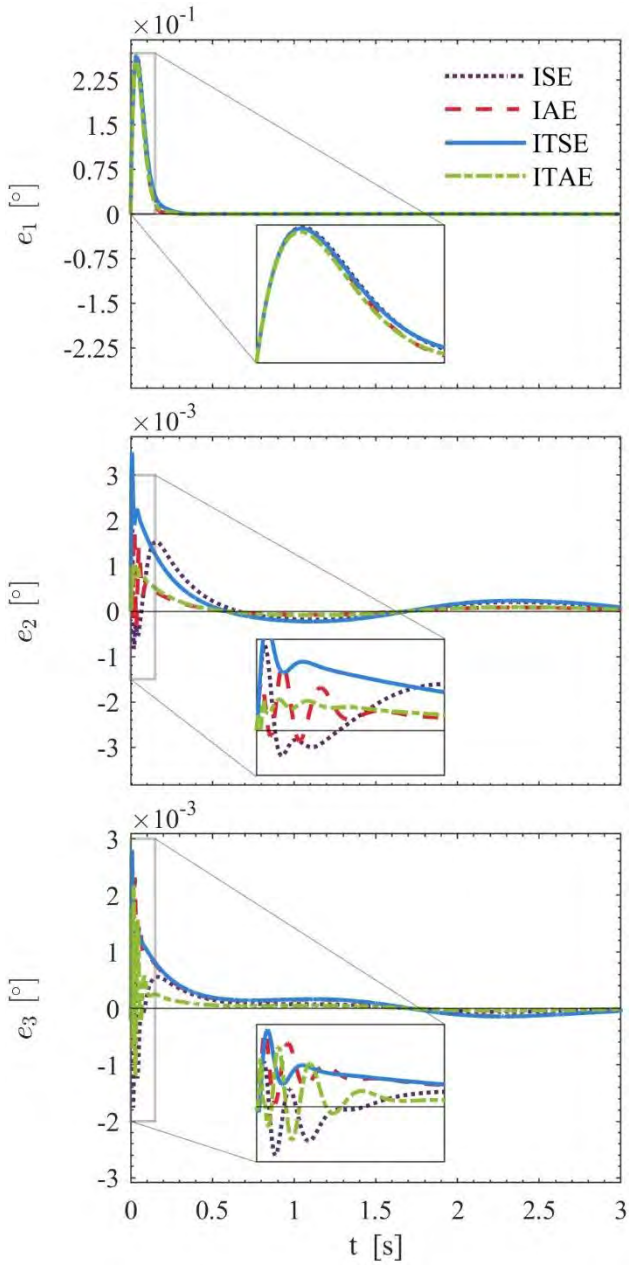


Figure 4. Joint position error at the 10-th iteration per joint comparison between ISE, IAE, ITSE and ITAE performance criteria.

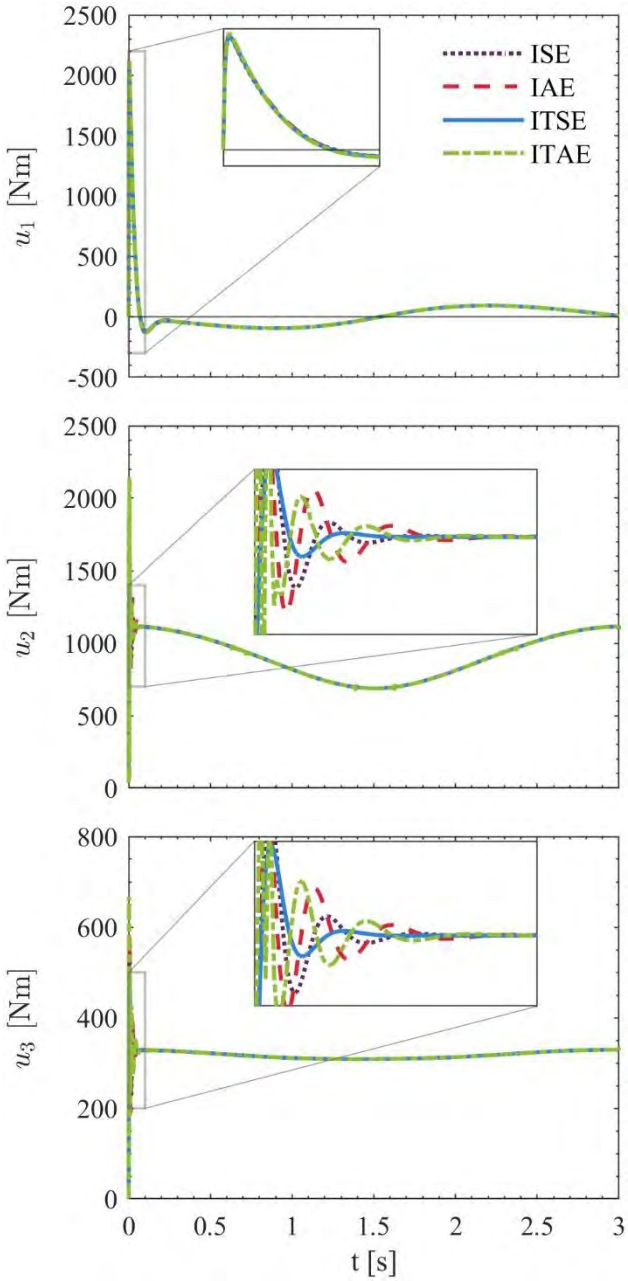


Figure 5. Joint control effort at the 10-th iteration per joint comparison between ISE, IAE, ITSE and ITAE performance criteria.

For joint 1, it is the same as in Figure 3: there are negligible differences among the selected criteria in performance. Joint 2 shows clearly that ITSE has the largest maximum absolute value among all selected criteria, and error tends to settle at zero for all criteria. Joint 3 error e3 again illustrates the same grouped distinction as in Figure 3.

Figure 5 shows joint control effort at the 10th iteration, and the most important conclusion is that none of the three joints enter torque saturation across all performance criteria.

4. CONCLUSION

This paper presented a control scheme combining Fractional-order Iterative Learning Control with a position PD controller for trajectory tracking of the LOLA15 6-DOF articulated robot. The central theme of this simulation study was the evaluation of performance across different criteria for PSO-based tuning of control parameters in the FOILC control law. Simulation results demonstrated that the proposed tuning procedure for the FOILC controller successfully improves tracking performance over successive iterations, and that, for FOILC parameter tuning of the robot LOLA15, the best criterion among the selected ones is ITAE, as it shows stable convergence and the best overall performance. The results show that the choice of optimization criterion affects both convergence behavior

and final tracking accuracy, and that this should be considered in PSO-based FOILC tuning. Future work will include experimental validation of the physical manipulator further to confirm the effectiveness of the proposed control approach.

Acknowledgment

This research has been supported by the research grants of the Serbian Ministry of Science, Technological Development, and Innovations, Grant No. 51-03-34/2026-03/ 200105 from 05.02.2026. and Grant No. 451-03-33/2026-03/200066.

References

[1] SPONG, M. W.: An historical perspective on the control of robotic manipulators, Annual Review of Control, Robotics, and Autonomous Systems, 5(1) (2022), 1-31.

[2] KELLY, R., DAVILA, V. S., LORÍA, A.: Control of Robot Manipulators in Joint Space, Advanced Textbooks in Control and Signal Processing, Springer, London, 2005.

[3] FLIESS, M., JOIN, C.: Model-free control. International journal of control, 86(12) (2013), 2228-2252.

[4] ARIMOTO, S., KAWAMURA, S., MIYAZAKI, F.: Bettering operation of robots by learning, Journal of Robotic Systems, 1(2) (1984), 123-140.

[5] MOORE, K. L.: Iterative Learning Control for Deterministic Systems, Springer-Verlag, London, 1993.

[6] ARIMOTO, S.: Iterative Learning Control, Springer US, Boston, 1998.

[7] LONGMAN, R. W.: Iterative learning control and repetitive control for engineering practice, International Journal of Control, 73(10) (2000), 930–954.

[8] MILLER, K. S., ROSS, B.: An introduction to the fractional calculus and fractional differential equations. New York: John-Wiley and Sons. Inc., 1993.

[9] LI, Y., CHEN, Y., AHN, H.-S., TIAN, G.: A Survey on Fractional-Order Iterative Learning Control, J Optim Theory Appl, 156(1) (2013) 127–140.

[10] SABATIER, J., LANUSSE, P., MELCHIOR, P., OUSTALOUP, A.: Fractional Order Differentiation and Robust Control Design: CRONE, H-infinity and Motion Control, Intelligent Systems, Control and Automation: Science and Engineering, 77, Springer Netherlands, Dordrecht, 2015.

[11] PRADHAN, R., MAJHI, S. K., PRADHAN, J. K., PATI, B. B.: Optimal fractional order PID controller design using Ant Lion Optimizer, Ain Shams Engineering Journal, 11(2) (2020), 281–291.

[12] VANCHINATHAN, K., SELVAGANESAN, N.: Adaptive fractional order PID controller tuning for brushless DC motor using Artificial Bee Colony algorithm, Results in Control and Optimization, 4 (2021), 100032.

[13] GHASEMI, I., RANJBAR NOEI, A., SADATI, J.: Intelligent fractional order iterative learning control using feedback linearization for a single-link robot, IIUMEJ, 18(1) (2017), 155–176.

[14] ŽIVKOVIĆ, N., LAZAREVIĆ, M., VIDAKOVIĆ, J.: Fractional order Iterative learning controller with PSO-based parameters tuning for robotic systems, Proceedings of the 11th International Scientific Conference on Defensive Technologies - OTEH 2024, 307-312.

[15] ŽIVKOVIĆ, N., LAZAREVIĆ, M., VIDAKOVIĆ, J.: Fractional-order iterative learning control enhanced intelligent PID for articulated robots, International Journal of Control, Automation and Systems, 23(7) (2025), 2036-2045.

[16] ŽIVKOVIĆ, N., LAZAREVIĆ, M., VIDAKOVIĆ, J.: Assessment of fractional order impact on performance of fractional ILC controller for upper limb exoskeleton, 16th International Conference On Accomplishments In Mechanical And Industrial Engineering, DEMI 2023, Banja Luka 1-2 June, pp. 333–337, 2023.

[17] EBERHART, R., KENNEDY, J.: Particle swarm optimization, Proceedings of the IEEE international conference on neural networks, (1995) 1942–1948.

[18] RAPAIĆ, M. R., KANOVIĆ, Ž.: Time-varying PSO – convergence analysis, convergence-related parameterization and new parameter adjustment schemes, Information Processing Letters, 109(11) (2009), 548–552.

[19] KVRGIC, V., VIDAKOVIC, J.: Efficient method for robot forward dynamics computation, Mechanism and Machine Theory, 145, (2020), 103680.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

METHOD AND SYSTEM FOR THROUGH-THE-WALL RADAR IMAGING

MLADEN ŽIVKOVIĆ
Military Academy, Belgrade, Serbia, zm215039p@student.etf.bg.ac.rs

Abstract: *Through-the-wall radar imaging is a challenging problem in integrated sensor systems due to strong reflections from walls that obscure weak target responses. This paper presents a method and system for through-the-wall imaging based on the processing of numerically generated scattering data, without requiring prior knowledge of wall characteristics or reference measurements. The proposed approach focuses on enhancing target visibility and improving image quality in complex and obstructed environments. The proposed method and system are intended for integration into advanced sensor and robotic platforms requiring reliable situational awareness. Numerical results demonstrate effective suppression of unwanted reflections and improved target localization performance. The approach is evaluated in comparison with conventional imaging techniques, showing competitive or improved results in terms of robustness and detection capability. Reliable detection of targets behind walls is achieved, with potential applications in surveillance, search and rescue, and autonomous sensing systems.*

Keywords: *Through-the-wall radar imaging, wall clutter suppression, target detection, integrated sensor systems, robotic sensing.*

1. INTRODUCTION

Through-the-wall radar imaging (TWRI) represents an important sensing technology for the detection and localization of targets in visually obscured environments. In many practical scenarios, such as surveillance, search and rescue operations, disaster response, and autonomous robotic inspection, optical sensors and conventional imaging systems are ineffective because the region of interest is hidden behind walls or other obstacles [1]-[2]. Radar systems, due to their ability to penetrate dielectric materials and operate under poor visibility conditions, provide a promising solution for obtaining situational awareness in such environments.

Despite its potential, through-the-wall imaging remains a highly challenging problem. The electromagnetic wave transmitted by the radar is strongly affected by the wall, resulting in reflections, attenuation, phase distortion, multipath propagation, and clutter. In many cases, the wall response is significantly stronger than the response of the target behind it. As a consequence, weak target signatures can be masked by dominant wall reflections, which reduces the quality of the reconstructed image and degrades target localization accuracy. This problem becomes even more difficult when the wall parameters are unknown or when reference measurements of the empty scene are not available.

A common approach for wall clutter mitigation is background subtraction, where a reference measurement containing only the wall response is subtracted from the measurement containing the target [3]-[4]. However, such an approach requires access to a target-free scene, which is often unrealistic in real applications. In addition, small changes in the environment, antenna position, or wall properties can reduce the effectiveness of reference-based methods. Therefore, there is a need for imaging techniques that can suppress unwanted wall reflections directly from measured radar data, without requiring prior knowledge of the wall structure or reference measurements.

Recent research in radar sensing has also shown the importance of robust signal processing when radar systems are used in complex and obstructed environments. For example, radar mounted on unmanned aerial vehicles has been investigated for detecting and monitoring people in collapsed buildings, where strong reflections from static objects and platform motion must be compensated before weak human-related signals can be extracted [5]. Such systems demonstrate that reliable detection in difficult environments depends not only on the radar hardware, but also on advanced processing methods capable of suppressing unwanted responses and enhancing the desired target information.

In this paper, a method and system for through-the-wall radar imaging are proposed based on the processing of numerically generated scattering data. The main objective is to improve target visibility and localization performance in the presence of strong wall reflections and clutter. Unlike conventional approaches that rely on reference measurements or known wall parameters, the proposed method operates directly on the available measured data. By applying data-driven processing, the dominant unwanted components are reduced, while the weaker target-related responses are enhanced.

The proposed approach is intended for integration into advanced sensor and robotic platforms that require reliable situational awareness in obstructed environments. Such platforms may include mobile robots, unmanned ground vehicles, or aerial systems used in surveillance, emergency response, and search and rescue missions.

The performance of the proposed method is evaluated

Corresponding author: M. Živković;
E-mail: zm215039p@student.etf.bg.ac.rs
DOI: 10.5937/OTEH26065Z

using numerical data. The obtained results are compared with conventional imaging techniques in order to assess target localization accuracy, robustness, and clutter suppression capability. The results demonstrate that the proposed method can effectively suppress dominant wall reflections and improve the visibility of targets behind walls.

2. THROUGH-THE-WALL RADAR IMAGING SYSTEM AND METHOD

The proposed through-the-wall radar imaging system consists of a stationary linear antenna array, a measurement unit for acquiring complex scattering parameters, and a signal-processing module for wall-clutter suppression and image reconstruction (Figure 1).

A linear array containing M antenna elements is positioned in front of the wall, as illustrated in Figure 2. Each antenna element is successively used for signal transmission and reception. Consequently, the complete set of transmit–receive measurements forms the complex scattering matrix:

$$\mathbf{S} \in \mathrm{C}^{M\times M}. \tag{1}$$

The measured scattering matrix can be represented as

$$\mathbf{S} = \mathbf{S}_{\text{wall}} + \mathbf{S}_{\text{target}} + \mathbf{N}, \tag{2}$$

where $\mathbf{S}_{\text{wall}}$ represents the dominant wall reflection, $\mathbf{S}_{\text{target}}$ contains the target responses, and $\mathbf{N}$ denotes measurement noise and other unwanted components.

2.1. Signal Preprocessing

The wall reflection is generally considerably stronger than the field scattered from a target. Therefore, direct image reconstruction from the original scattering matrix may result in strong wall-related artifacts.

As the first preprocessing step, the measured matrix is centered column-wise. The mean column vector is calculated as

$$\bar{\mathbf{s}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{s}_m, \tag{3}$$

where $\mathbf{s}_m$ denotes the mth column of $\mathbf{S}$. The centered scattering matrix is then obtained as

$$\mathbf{S}_c = \mathbf{S} - \bar{\mathbf{s}}\mathbf{1}^T, \tag{4}$$

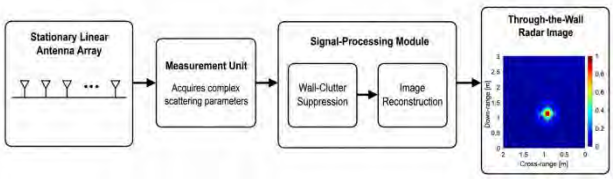


Figure 1. Proposed through-the-wall radar imaging system

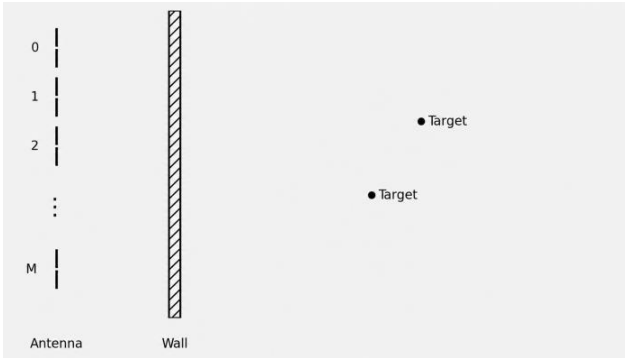


Figure 2. Geometry of the considered through-the-wall radar imaging system

where **1** is an all-ones vector of length M . This operation reduces the common-mode component present in the measured signals.

2.2. SVD-Based Wall-Clutter Suppression

Singular value decomposition is applied to the centered scattering matrix:

$$\mathbf{S}_c = \boldsymbol{U\Sigma V}^H, \tag{5}$$

where **U** and **V** contain the left and right singular vectors, respectively, while

$$\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_M). \tag{6}$$

contains the singular values arranged in descending order,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0. \tag{7}$$

The dominant wall reflection is assumed to occupy a low-rank subspace associated with the largest singular values. To determine the wall-subspace dimension, the singular values are first normalized:

$$\sigma_{norm,k} = \frac{\sigma_k}{\max\limits_i(\sigma_i)}. \tag{8}$$

The rank estimate is obtained by applying Otsu's threshold γ_{Otsu} to the normalized singular-value spectrum:

$$r = max\{k: \sigma_{norm,k} \geq \gamma_{Otsu}\}. \tag{9}$$

The low-rank wall component is reconstructed according to:

$$\mathbf{S}_{\text{wall}} = \mathbf{U}_{(:,1:r)}\boldsymbol{\Sigma}_{(1:r,1:r)}\mathbf{V}^H_{(:,1:r)}. \tag{10}$$

The sparse residual containing the target response is obtained by subtracting the estimated wall component:

$$\mathbf{S}_{\text{sparse}} = \mathbf{S}_c - \mathbf{S}_{\text{wall}}. \tag{11}$$

2.3. Regularized Image Reconstruction

The sparse scattering matrix is decomposed again using SVD:

$$\mathbf{S}_{\text{sparse}} = \mathbf{U}_t\boldsymbol{\Sigma}_t\mathbf{V}_t^H. \tag{12}$$

For every point $\mathbf{p} = (x_i, y_j)$ within the region of interest, the steering vector is calculated as

$$\mathbf{a}(\mathbf{p}) = \begin{bmatrix} \exp\left(-j\frac{2\pi f}{c}d_1(\mathbf{p})\right) \\ \exp\left(-j\frac{2\pi f}{c}d_2(\mathbf{p})\right) \\ \vdots \\ \exp\left(-j\frac{2\pi f}{c}d_M(\mathbf{p})\right) \end{bmatrix}, \quad (13)$$

where

$$d_m(\mathbf{p}) = \sqrt{\{(x_i - x_m)^2 + y_j^2\}} \quad (14)$$

is the distance between the mth antenna element and the considered image point, while c is the speed of light. The regularization parameter is determined from the singular values and the statistics of the wall-suppressed residual (sparse) matrix:

$$\alpha = \kappa \frac{\max(\boldsymbol{\sigma}_t)}{1 + \operatorname{std}(\mathbf{S}_{\text{sparse}})} \quad (15)$$

where

$$\kappa = 0.01 \frac{\operatorname{std}(\boldsymbol{\sigma}_t)}{\operatorname{mean}(\boldsymbol{\sigma}_t)} \quad (16)$$

For each sampling point $\mathbf{p}$, the regularized solution vector is calculated using a truncated Tikhonov filter:

$$\mathbf{f}(\mathbf{p}) = \sum_{k=1}^{r} \frac{\sigma_{t,k}}{\sigma_{t,k}^2 + \alpha^2} \left(\mathbf{u}_{t,k}^H \mathbf{a}(\mathbf{p})\right) \mathbf{v}_{t,k}. \quad (17)$$

Finally, the image intensity is defined as

$$L(\mathbf{p}) = \frac{1}{\|\mathbf{f}(\mathbf{p})\|_2^2} \quad (18)$$

3. NUMERICAL RESULTS AND METHOD COMPARISON

The performance of the proposed through-the-wall radar imaging system was evaluated using scattering parameters numerically generated in WIPL-D Pro, a full-wave three-dimensional electromagnetic solver based on the Method of Moments [7]. The objective was to determine whether the proposed processing method can suppress the dominant wall response and preserve the weaker fields scattered from targets located behind the wall.

The proposed method was compared with the Multiple Signal Classification (MUSIC) method [3] and the Linear Sampling Method (LSM) [6]. All methods were evaluated using the same antenna geometry, wall configuration, region of interest, and spatial discretization.

3.1. Imaging Scenario

The considered TWRI configuration consists of a uniform linear antenna array placed at a distance of 0.3 m from the front surface of the wall. The antenna array consists of M = 21 dipoles operating at a frequency of f = 1 GHz. The region of interest (ROI) is discretized into a rectangular grid spanning x ∈ [0, 2] meters in the cross-range and down-range directions y ∈ [0, 3] meters, with 100 × 100 pixel resolution.

The wall is placed in front of the array, while the targets are positioned behind it, as illustrated in Figure 2. The wall dimensions are 2×2×0.2 m, where 0.2 m represents the wall thickness. The wall is modeled as a homogeneous dielectric medium with a relative permittivity of $\varepsilon_r = 6$.

The targets are modeled as perfectly electrically conducting (PEC) objects with the shape of small human figures. Two target configurations are considered. In the single-target scenario, the target is located at (0.75, 1, 0) m. In the multi-target scenario, three targets are positioned at (1, 1.5, 0) m, (1.75, 2.5, 0) m, and (2, 2 ,0) m, respectively. These configurations allow the localization performance and robustness of the proposed method to be evaluated under different scene complexities.

For MUSIC and LSM, the background-subtracted scattering matrix is calculated as

$$\mathbf{S}_\Delta = \mathbf{S}_{target} - \mathbf{S}_{bg}, \quad (19)$$

where $\mathbf{S}_{target}$ denotes the scattering matrix measured with the targets present and $\mathbf{S}_{bg}$ represents the corresponding background measurement without targets.

In contrast, the proposed method operates directly on the measured matrix $\mathbf{S}$, without requiring a background measurement or prior knowledge of the electromagnetic properties of the wall.

3.2. Reference Imaging Methods

The LSM reconstructs an indicator function by solving a regularized inverse problem for each sampling point within the imaging region. The method uses the background-subtracted matrix $\mathbf{S}_\Delta$ as the operator kernel [6].

The norm of the regularized solution is used to define the imaging indicator. High indicator values correspond to positions at which a target is likely to be present.

The MUSIC method estimates target positions by projecting the steering vector onto the noise subspace. For each point $\mathbf{p}$ within the region of interest, the MUSIC pseudo-spectrum is calculated as

$$P_{\text{MUSIC}}(\mathbf{p}) = \frac{1}{\mathbf{a}^H(\mathbf{p})\mathbf{E}n\mathbf{E}_n^H\mathbf{a}(\mathbf{p})}, \quad (20)$$

where $\mathbf{E}n$ contains the eigenvectors associated with the noise subspace. Large values of $P_{\text{MUSIC}}(\mathbf{p})$ indicate possible target locations [3].

3.3. Single-Target Results

Figure 3 compares the reconstructed images obtained using LSM, MUSIC, and the proposed method for the single-target scenario. The actual positions and dimensions of the targets are indicated by black squares.

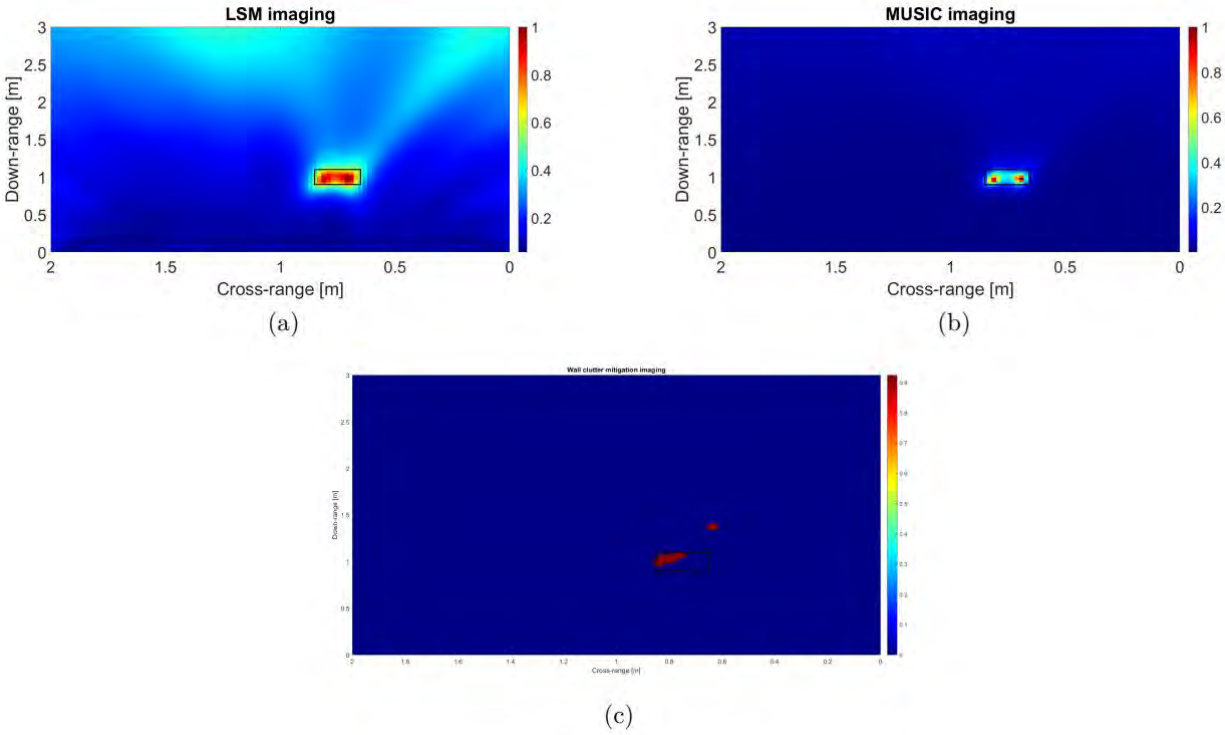


Figure 3. Comparison of reconstructed images for the single-target scenario: (a) LSM, (b) MUSIC, and (c) proposed method.

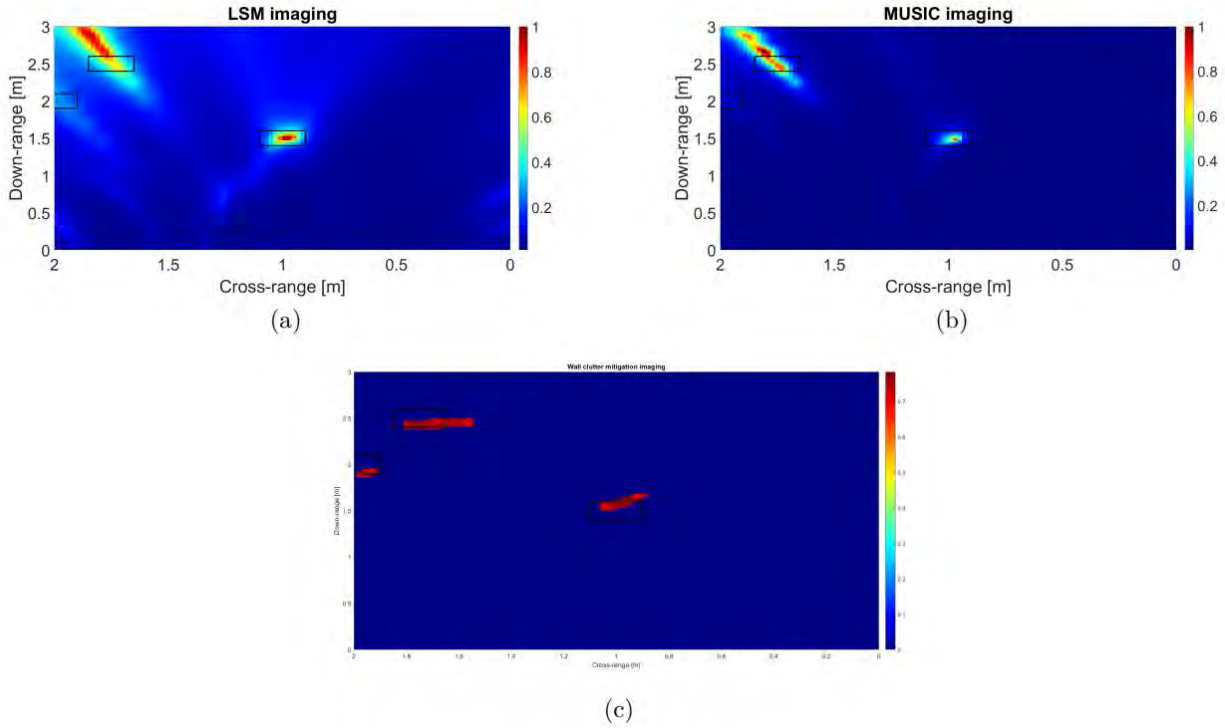


Figure 4. Comparison of reconstructed images for the multi-target scenario: (a) LSM, (b) MUSIC, and (c) proposed method.

The LSM reconstruction indicates the approximate target position, but the target response is broadened and accompanied by an elevated background level. These effects reduce image contrast and may make precise target localization difficult.

The MUSIC method produces a more compact response and a lower background level. However, the reconstructed target region contains a split peak, indicating that the target response is distributed between two closely spaced image maxima.

The proposed method significantly attenuates the dominant wall reflection and produces a clearly visible response in the vicinity of the actual target position. Minor isolated artifacts may remain outside the main target region, but their intensity is considerably lower than the dominant target response. The results demonstrate that the wall clutter can be suppressed without using a reference measurement.

3.4. Multi-Target Results

Figure 4 presents the reconstructed images for the scenario containing three targets behind the wall.

The LSM reconstruction contains broadened target responses, wall-related artifacts, and sidelobe-like structures. Although the regions associated with the three targets can still be observed, the elevated background level and spatial spreading make their separation and precise localization more difficult.

MUSIC provides more compact responses and a cleaner background, allowing the three target regions to be distinguished more clearly. Nevertheless, residual energy and weak leakage remain around the dominant responses.

The proposed method reveals scattering regions associated with all three targets while substantially suppressing the direct wall response. However, compared with the single-target case, additional spurious responses and spatial smearing are present. These effects reduce the separability of closely spaced targets and may introduce localization errors.

The reduced robustness in the three-target scenario indicates that decomposition of the measured matrix into a low-rank wall component and a sparse target component becomes more challenging as the number of scatterers increases. Interactions between target responses, residual wall components, and multipath effects can further complicate the reconstruction process.

4. CONCLUSION

This paper presented a through-the-wall radar imaging method based on SVD-based wall-clutter suppression. The proposed approach operates directly on the available scattering matrix and does not require a background measurement or prior knowledge of the wall parameters.

Numerical results obtained from scattering parameters generated in WIPL-D demonstrated that the method substantially reduces the dominant wall response while preserving target-related scattering information. In the single-target scenario, the proposed method produced a focused target response with a low background level and performance comparable to MUSIC and LSM, although both reference methods relied on background-subtracted data.

In the three-target scenario, the proposed method continued to indicate the target regions, but residual artifacts and spatial smearing reduced target separability and localization accuracy. This result shows that the low-rank wall and sparse target decomposition becomes more challenging as the number of scatterers and scene complexity increase.

Future work will focus on improving multi-target robustness through multifrequency processing, adaptive wall-subspace rank selection, enhanced sparse regularization, and more accurate propagation models that account for wall refraction, dispersion, and multipath effects.

References

[1] YOON,Y.,S., AMIN,M.,G.: *Spatial Filtering for Wall-Clutter Mitigation in Through-the-Wall Radar Imaging*, IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 9, pp. 3192-3208, Sept. 2009, doi: 10.1109/TGRS.2009.2019728.

[2] AMIN,M.,G.: *Through-the-Wall Radar Imaging*, CRC Press, 2017.

[3] YOON,Y.,S., AMIN,M.,G.: *High-Resolution Through- the-Wall Radar Imaging Using Beamspace MUSIC*, IEEE Transactions on Antennas and Propagation, vol. 56, no. 6, pp. 1763-1774, June 2008, doi: 10.1109/TAP.2008.923336.

[4] FIORANELLI,F., SALOUS,S., NDIP,I., RAIMUNDO,X.: *Through-The-Wall Detection With Gated FMCW Signals Using Optimized Patch-Like and Vivaldi Antennas*, IEEE Transactions on Antennas and Propagation, vol. 63, no. 3, pp. 1106-1117, March 2015, doi: 10.1109/TAP.2015.2389793.

[5] STOCKEL,P., WALLRATH,R., HERSCHEL,R., POHL,N.: *Detection and monitoring of people in collapsed buildings using a rotating radar on a UAV*, IEEE Transactions on Radar Systems, vol. 2, pp. 13–23, 2024, doi: 10.1109/TRS.2023.3342368.

[6] CATAPANO,I., CROCCO,L.: *A qualitative inverse scattering method for through-the-wall imaging*, IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 4, pp. 685–689, Oct. 2010, doi: 10.1109/LGRS.2010.2045473.

[7] WIPL-D Team, *WIPL-D Pro v21.0: 3D Electromagnetic Solver, Professional Edition - User's Manual*. Belgrade, Serbia: WIPL-D d.o.o., 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AUTOMATED IMAGE CAPTIONING WITH A RESNET-LSTM ENCODER-DECODER FOR DISTRIBUTED SURVEILLANCE SCENES

ANĐA DENIĆ
Faculty of Electronic Engineering, University of Niš, Niš, andja.denic@elfak.ni.ac.rs

SNEŽANA ZUROVAC
Military Technical Institute, Belgrade, zurovac@medianis.net

IVANA PETKOVIĆ
Military Technical Institute, Belgrade, ivana.petkovic.vti@gmail.com

Abstract: *Automatic image captioning is a fundamental artificial intelligence task at the intersection of computer vision and natural language processing. This paper presents a generative encoder-decoder model for automatic caption generation from images of military exercises. The model employs an ImageNet-pre-trained 152-layer Residual Network (ResNet-152) encoder to extract visual representations and a long short-term memory (LSTM) decoder to generate corresponding textual descriptions. The original model was trained on the Microsoft Common Objects in Context (MS COCO) dataset and subsequently fine-tuned on a domain-specific collection of military-exercise images. Five pseudo-reference captions per image were generated automatically using Florence-2 and used to supervise domain adaptation.*

Experimental evaluation was performed on a held-out military-domain test set containing 25 images. The fine-tuned model achieved Bilingual Evaluation Understudy scores of 70.87, 49.67, 33.61, and 21.94 for BLEU-1, BLEU-2, BLEU-3, and BLEU-4, respectively, together with a Recall-Oriented Understudy for Gisting Evaluation based on the longest common subsequence (ROUGE-L) score of 45.52. On the same military test set, the original MS COCO checkpoint achieved 52.00, 33.10, 23.00, 15.80, and 37.05, respectively. These results indicate that domain-specific fine-tuning improves agreement with the available military-domain reference captions.

However, because the test set is small and its reference captions were generated automatically rather than independently written by domain experts, the reported results should be interpreted as preliminary evidence rather than a definitive estimate of performance on the broader military-image domain. The proposed architecture may provide a basis for future investigation of onboard semantic reporting and bandwidth-efficient communication on unmanned platforms, while its practical contribution can be further experimentally validated through the evaluation of execution speed, energy consumption, and communication savings on a real unmanned aerial vehicle.

Keywords: *image captioning, LSTM, ResNet, encoder-decoder, domain-specific dataset, military-exercise*

1. INTRODUCTION

Automatic image captioning addresses one of the central challenges in machine learning: converting visual perception into structured natural language. Unlike image classification or object detection, where the output is usually a fixed label or a set of localized categories, caption generation requires a model to identify objects, infer their attributes and relations, and express the resulting scene understanding as a grammatically coherent sentence. This makes image captioning a representative vision-language problem at the intersection of computer vision and natural language processing.

The practical relevance of automatic image captioning is particularly strong in scenarios where large volumes of visual data must be interpreted and communicated efficiently. In surveillance, reconnaissance, digital archiving, assistive technologies, and autonomous robotic systems, raw image and video streams can contain substantially more data than can be transmitted, stored, or inspected in real time. A compact textual description may provide a semantic summary of a scene and support filtering, indexing, prioritization, and operator decision-making. For distributed systems, including unmanned platforms, onboard caption generation could potentially reduce the amount of visual data that must be transmitted to a central station. Nevertheless, the present study evaluates only captioning accuracy on an offline image dataset. It does not experimentally evaluate onboard inference, real-time processing, energy consumption, communication latency, or operation on an actual unmanned platform.

In this paper, we study automatic caption generation for images of military exercises. This domain introduces a meaningful distribution shift with respect to general-purpose image-captioning datasets because the scenes may contain specialized equipment, formations, activities, uniforms, vehicles, and operational contexts that are underrepresented in broad collections of everyday

Corresponding author: A. Denić;
E-mail: andja.denic@elfak.ni.ac.rs
DOI: 10.5937/OTEH26066D

imagery. The original model was trained on the Microsoft Common Objects in Context (MS COCO) dataset, which provides useful general visual and linguistic regularities but does not adequately represent all concepts found in military-training environments. Domain adaptation is therefore investigated as a means of improving the relevance of generated descriptions for the target imagery.

The proposed approach follows an encoder-decoder paradigm. A 152-layer Residual Network (ResNet-152) acts as the visual encoder and transforms an input image into a fixed-dimensional representation. A long short-term memory (LSTM) network then generates the caption sequentially, one word at a time. The model is first trained on MS COCO and subsequently fine-tuned on automatically annotated military-exercise images. The principal contribution is an end-to-end domain-adaptation pipeline that connects target-domain data collection, automatic annotation, deterministic image preprocessing, image-level dataset partitioning, parameter-efficient fine-tuning, and comparative evaluation of the original and adapted checkpoints.

This architecture offers a favorable balance between modeling capacity and deployment practicality. While large vision-language foundation models have recently achieved impressive performance, compact encoder-decoder systems remain attractive for edge scenarios where computational resources, latency, power consumption, and communication bandwidth are constrained. In such settings, a ResNet-LSTM model can provide interpretable and efficient caption generation without requiring continuous transmission of high-resolution visual streams.

The main contribution of this work is an end-to-end captioning pipeline adapted to military-exercise imagery. The paper demonstrates how a general image-captioning model can be transferred from MS COCO to a specialized operational domain, evaluates the impact of fine-tuning using standard captioning metrics, and discusses the suitability of the resulting system for distributed surveillance and drone-based deployment. Experimental results indicate that domain-specific fine-tuning improves the quality and relevance of generated captions compared with the base model trained only on generic data.

2. THEORETICAL BACKGROUND AND RELATED WORK

The present study extends the encoder-decoder framework examined in the preceding work. A central instance of this framework is the neural image-caption generator introduced by Vinyals et al. [1], in which a convolutional neural network (CNN) encodes an image and a recurrent neural network (RNN) estimates the probability of a caption conditioned on the resulting visual representation.

Let I denote an input image and Y its caption of length T. The model factorizes the conditional probability of the complete caption autoregressively as

$$P_\theta(Y \mid I) = \prod_{t=1}^{T} P_\theta(y_t \mid I, y_1, y_2, \dots, y_{t-1}) \qquad (1)$$

where θ contains the trainable parameters of the visual encoder, word-embedding layer, recurrent decoder, and output classifier. Learning is performed by maximizing the conditional likelihood of the reference captions, or equivalently by minimizing token-level cross-entropy. This probabilistic formulation provides a direct connection between visual representation learning and sequence modeling and is the theoretical basis of both the original model and its domain-specific adaptation considered here.

The visual encoder used in this work is based on ResNet-152. Residual networks introduce identity shortcuts that allow a stack of nonlinear layers to learn a residual mapping rather than an unconstrained direct mapping, thereby facilitating the optimization of very deep convolutional architectures [2]. In the adopted implementation, the classification layer of a ResNet-152 model pre-trained on ImageNet is removed. Global image features produced by the convolutional backbone are transformed by an additional linear layer and batch-normalization layer into a fixed-dimensional context vector. Transfer learning is particularly important in the present setting because a comparatively small collection of military-exercise images is insufficient for learning a deep visual representation from random initialization.

The decoder is a long short-term memory network. LSTM units were designed to model long-range dependencies through a recurrent cell state regulated by input, forget, and output gates [3]. For image captioning, the encoded image is supplied to the recurrent model before the textual sequence, after which the decoder predicts one word at a time. The probability distribution over the vocabulary at time step t depends on the current LSTM hidden state, which summarizes the image representation and the previously observed or generated words. During training, reference tokens are provided as preceding inputs, whereas inference is autoregressive and uses the model's own previous predictions.

MS COCO dataset provides a broad source domain for learning general visual and linguistic regularities because it contains complex everyday scenes paired with multiple human-written captions [4]. Nevertheless, its content distribution differs from that of military-training imagery. Uniforms, formations, specialized vehicles, aircraft, vessels, weapons, field exercises, and operational activities may occur rarely or under different contextual priors in a general-purpose dataset. As a result, a model trained only on MS COCO can produce grammatically plausible captions while omitting domain-relevant details or selecting vocabulary that is semantically inappropriate for the target environment. The problem addressed in this paper is therefore one of domain adaptation: knowledge acquired from a large generic dataset is retained, while the trainable components are updated using target-domain image-caption pairs.

The broader development of vision-language modeling has introduced attention mechanisms, Transformer architectures, and large prompt-conditioned foundation models. Visual attention allows a decoder to select spatially relevant image regions while generating

individual words [5], whereas Transformer-based models replace recurrence with attention-based sequence processing [6]. Florence-2 represents a more recent unified approach: it treats captioning, detection, grounding, and related tasks as prompt-conditioned sequence generation within a single vision-language model [7]. In the present pipeline, Florence-2 is not the final deployed captioning model and its parameters are not fine-tuned. Instead, the pre-trained Florence-2 model is used as an automatic annotation system that generates pseudo-reference captions for the military-exercise images. These annotations subsequently supervise the fine-tuning of the more compact ResNet-LSTM model.

Domain-specific captioning has also been investigated for remote-sensing imagery, where the distribution of viewpoints, spatial structures, and semantic categories differs substantially from ordinary ground-level photography [8]. This line of work illustrates the general principle relevant to defence-oriented data: a captioning architecture may be transferable, but its vocabulary and conditional output distribution must be adapted to the target domain. For surveillance and drone-based platforms, compact encoder-decoder systems remain attractive because they can convert high-bandwidth visual observations into short semantic reports locally. Such use does not eliminate the need for human verification, particularly in safety-critical or operational contexts, but it can support filtering, indexing, prioritization, and bandwidth-efficient communication.

3. METHODOLOGY AND DATA

3.1. Dataset

The target-domain dataset was constructed from publicly available photographs of military exercises published on the official website of the Ministry of Defence of the Republic of Serbia. The images depict military personnel, formations, training procedures, helicopters, boats, ground vehicles, equipment, and outdoor exercise environments that are not adequately represented in general-purpose captioning datasets. The complete collection contains 167 distinct images. Of these, 142 images are used for training and validation, while 25 images form a held-out military-domain test set.

Because manually producing several independent descriptions for every image is time-consuming, the images were automatically annotated using the open-source Florence-2 model, specifically microsoft/Florence-2-base-ft. Annotation was performed using a graphics processing unit (GPU), and five pseudo-reference captions were generated for each image. These captions provide scalable supervision for fine-tuning the smaller ResNet-LSTM model, but they should not be treated as equivalent to independently written expert annotations.

The quality of the automatically generated captions directly affects both model training and evaluation. Florence-2 may omit relevant objects or actions, use overly general terminology, repeat similar sentence structures, or occasionally introduce an object that is not present in the image. Such errors act as label noise during fine-tuning and can encourage the adapted model to reproduce the same omissions, biases, or hallucinations. Moreover, because the evaluation references were generated by the same automatic annotation procedure, BLEU and ROUGE-L primarily measure agreement with Florence-2 wording and may not fully reflect factual correctness or agreement with a military-domain expert. Five captions produced by one model are also likely to be linguistically correlated and therefore do not provide the same diversity as five descriptions written independently by human annotators.

The training-validation split is performed at the image level rather than at the image-caption-pair level. Unique image identifiers are sorted, shuffled using a fixed pseudorandom seed, and partitioned using a validation ratio of 0.15. This produces 121 training images with 605 captions and 21 validation images with 105 captions. Because all captions belonging to one image remain in the same subset, direct image leakage between training and validation data is prevented. The independent test collection contains 25 images and 125 automatically generated reference captions and is used to compare the original and fine-tuned checkpoints.

The limited size of the target dataset is an important limitation of the study. In particular, results calculated on only 25 test images can be sensitive to the presence or absence of a small number of scenes, objects, and environmental conditions. Consequently, the reported metric improvements may not generalize to the full diversity of military imagery, including different exercises, equipment categories, seasons, weather conditions, illumination levels, viewpoints, sensors, and geographic environments. The current evaluation should therefore be regarded as an initial controlled comparison on the available test collection. A larger independently collected test set, preferably with captions reviewed or written by domain experts, is required for a more reliable estimate of generalization.

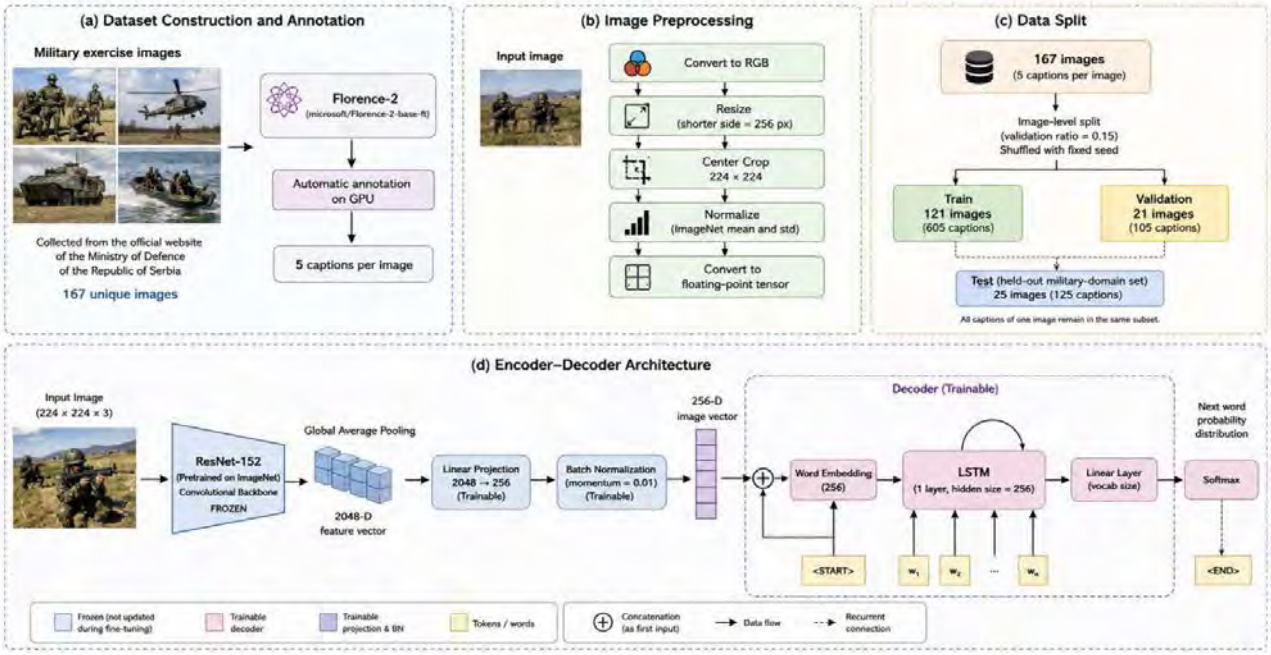


Figure 1. Overview of the methodology: (a) dataset construction and automatic annotation with Florence-2; (b) image preprocessing pipeline; (c) train/validation/test split performed at the image level; (d) encoder-decoder architecture with a frozen ResNet-152 encoder and trainable LSTM decoder

Figure 1 presents an overview of the complete methodology used in this study, from dataset construction and preprocessing to model training and inference. Before being passed to the ResNet encoder, each image is opened with the Python Imaging Library, converted to RGB, resized so that the shorter side equals 256 pixels, center-cropped to 224×224 pixels, and converted to a floating-point tensor. Channel-wise normalization uses the ImageNet statistics $\mu=(0.485, 0.456, 0.406)$ and $\sigma=(0.229, 0.224, 0.225)$, which is consistent with the pre-training of the ResNet backbone. The same deterministic transformation is used during fine-tuning and evaluation.

Because manually producing several independent captions for every image is costly, the images were automatically annotated using the open-source Florence-2 model (microsoft/Florence-2-base-ft). Annotation is performed on the GPU. Five captions have been collected for an image.

Fine-tuning performed the training-validation split at the image level rather than at the image-caption-pair level. Unique image identifiers are sorted, shuffled using a fixed pseudorandom seed, and partitioned using a validation ratio of 0.15. This yields 121 training images and 21 validation images. Because all captions associated with one image remain in the same subset, the split prevents direct image leakage between training and validation data. The held-out military-domain test collection is reserved for the comparison between the original and fine-tuned models.

3.2. Encoder-Decoder Architecture

The target model retains the architecture developed in the preceding study [1]. Its encoder is an ImageNet-pre-trained ResNet-152 with the original classification layer removed. After global average pooling, the convolutional backbone produces a 2,048-dimensional vector. A trainable linear projection maps this vector to a 256-dimensional joint representation, followed by batch normalization with momentum 0.01. All parameters in the ResNet convolutional backbone are frozen, whereas the projection and batch-normalization parameters remain trainable.

The decoder contains a word-embedding layer of dimension 256, a single-layer LSTM with a hidden-state dimension of 256, and a linear output layer whose dimension equals the vocabulary size. The image vector and word embeddings consequently occupy the same dimensional space. Following the Show and Tell formulation [1], the encoded image is inserted as the first input to the LSTM, before the <START> token and caption words. The final linear layer transforms each LSTM hidden state into unnormalized vocabulary scores, and a softmax interpretation yields the conditional probability of the next token.

Captions are converted to lowercase and punctuation is removed. The sequence is tokenized by whitespace, mapped to indices from the vocabulary inherited from the MS COCO model, and surrounded by <START> and <END> markers. Variable-length sequences are sorted within each mini-batch, padded, and packed for efficient recurrent processing. Tokens absent from the inherited vocabulary are excluded by the current preprocessing implementation rather than being used as explicit <UNK> targets. This choice preserves compatibility with the pre-trained decoder checkpoint but may remove specialized domain terminology; extending the vocabulary is therefore an important direction for subsequent work.

3.3. Training on MS COCO

The base model parameters are those obtained through training on MS COCO 2017. The source-domain training set contains more than 118,000 images, and the validation set contains 5,000 images, with approximately five human-written captions per image. The model uses a 256-dimensional image and word representation, a 256-dimensional LSTM state, and one recurrent layer. Training is conducted for 10 epochs with mini-batches of 64 examples and the Adam optimizer with a learning rate of 10^{-3}.

For a mini-batch, the encoder first produces image feature vectors, and the decoder generates vocabulary scores under teacher forcing.

The trainable parameters, denoted as θ, are the solutions to an optimization problem that maximizes the conditional likelihood of the reference captions given the input images.

The optimizer updates the complete decoder together with the encoder's linear projection and batch-normalization layer; the ResNet-152 convolutional backbone remains fixed. At the end of each epoch, the average validation loss is computed without gradient tracking. The encoder and decoder checkpoints associated with the lowest validation loss are retained for inference and evaluation. At inference time, the model uses greedy decoding: the most probable word is selected at each step until the <END> token is produced or a maximum caption length of 20 tokens is reached.

3.4. Domain-Specific Fine-Tuning

Fine-tuning starts from the saved encoder and decoder checkpoints of the MS COCO model; no component is reinitialized. The same Encoder CNN and Decoder RNN classes, vocabulary, image transformation, loss definition, and optimizer family are used to ensure architectural compatibility with the source model. The target-domain training and validation loaders use a batch size of 64.

The domain-adaptation stage is run for five epochs with Adam and a learning rate of 10^{-3}. Consequently, the frozen ResNet backbone preserves general ImageNet visual features, while the projection and language decoder adapt the joint image-text representation and conditional word distribution to the military-exercise domain. This parameter-efficient strategy reduces the computational cost and the risk of overfitting relative to updating all ResNet-152 layers on the limited target dataset.

Each epoch consists of a training phase and a validation phase. Whenever the validation loss improves, the fine-tuned state dictionaries overwrite the previous encoder and decoder checkpoints.

4. EXPERIMENTAL RESULTS

The experiments were conducted using an NVIDIA Quadro RTX 4000 graphics processing unit (GPU) with 8 GB of video memory, an eight-core central processing unit (CPU) supporting 16 processing threads, and 64 GB of random-access memory (RAM). The implementation was developed in Python using PyTorch. Torchvision provides the pre-trained ResNet-152 architecture and image transformations, while Florence-2 annotations are generated through the Hugging Face Transformers interface.

Caption quality is evaluated using the Bilingual Evaluation Understudy (BLEU) metric from BLEU-1 through BLEU-4 and the Recall-Oriented Understudy for Gisting Evaluation based on the longest common subsequence (ROUGE-L). BLEU measures modified n-gram precision with a brevity penalty [9], with larger n placing progressively greater emphasis on multiword correspondence. ROUGE-L measures sequence-level overlap using the longest common subsequence and therefore considers both shared content and word order [10]. All values in Table 1 are reported as percentages.

Table 1. Captioning performance on the base and military-domain evaluation sets (%)

Model / evaluation domain	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L
Base model / MS COCO	68.36	50.35	35.70	25.23	49.51
Original model / military test set	52.00	33.10	23.00	15.80	37.05
Fine-tuned model / military test set	**70.87**	**49.67**	**33.61**	**21.94**	**45.52**

When the original MS COCO checkpoint is applied directly to the military test set, all metrics decrease relative to its source-domain performance, confirming the presence of a distribution shift. Fine-tuning improves BLEU-1, BLEU-2, BLEU-3, BLEU-4, and ROUGE-L on the same military test set by 18.87, 16.57, 10.61, 6.14, and 8.47 percentage points, respectively. Because both checkpoints are evaluated using the same images and reference captions, this comparison provides evidence that adaptation improves agreement with the available target-domain annotations.

Nevertheless, the military test set contains only 25 images. Metric values calculated from such a small collection may have high sampling variability and can be disproportionately influenced by a few easy or difficult examples. No confidence intervals or repeated evaluations on independently sampled military test sets are available. The observed improvements are therefore valid for the evaluated collection but should not be interpreted as a precise estimate of performance across the broader military-image domain.

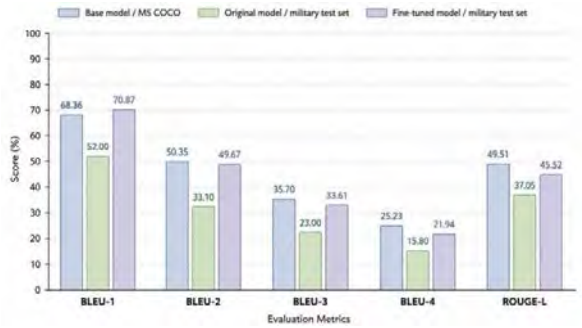


Figure 2. Captioning performance on the base (MS COCO) and military-domain evaluation sets. The fine-tuned model outperforms the original model on all metrics of the military test set

Figure 2 provides a visual comparison of the captioning performance achieved by the base and fine-tuned models on the MS COCO and military-domain evaluation datasets using the BLEU and ROUGE-L metrics. The scores obtained by the base model on MS COCO are included to document its source-domain performance, but they should not be interpreted as a controlled comparison with the military-domain scores because the reference captions and image distributions differ. The relevant evidence for the effectiveness of fine-tuning is the comparison of the two checkpoints on the same military test set.

Figure 3. Comparison of captions generated by the original model and the fine-tuned model for the first representative image from the military test set

Original model: “a man is sitting on a chair with a laptop”
Fine-tuned model: “two soldiers are sitting inside of a military vehicle”

Figure 4. Comparison of captions generated by the original model and the fine-tuned model for the second representative image from the military test set

Original model: “a group of people standing around a field of grass”

Fine-tuned model: “a group of soldiers in military uniforms are standing in a line”

Figure 5. Comparison of captions generated by the original model and the fine-tuned model for the second representative image from the military test set

Original model: “a group of people standing near a fire”

Fine-tuned model: “a group of soldiers walking through a forest”

Figure 6. Comparison of captions generated by the original model and the fine-tuned model for the second representative image from the military test set

Original model: “people standing beside a yellow truck”

Fine-tuned model: “a group of soldiers standing next to a tank”

Figures 3–6 illustrate both the benefits and the remaining limitations of domain-specific fine-tuning. In Figure 3, the original model focuses on familiar indoor objects and produces the caption “a man is sitting on a chair with a laptop,” whereas the fine-tuned model identifies the military context and describes two soldiers inside a military vehicle. Similarly, in Figure 4, the original model generates the overly general description “a group of people standing around a field of grass,” while the fine-tuned model more accurately identifies soldiers in military uniforms standing in a line.

The examples also reveal several characteristic errors. For the low-light firefighting scene in Figure 5, the caption “a group of people standing near a fire” is broadly related to the image but incomplete: it does not identify the people as military firefighters, does not describe their protective

equipment, and omits the active response to a forest fire.

Figure 6 demonstrates that domain adaptation can improve one part of a caption while introducing a different error. The original model describes the scene as “people standing beside a yellow truck,” which recognizes the presence of people and a large yellow vehicle but does not identify either the military personnel or the tracked construction machine. The fine-tuned model generates “a group of soldiers standing next to a tank.” This caption correctly introduces the military personnel but incorrectly classifies the tracked bulldozer as a tank. The error suggests that the adapted model may over-apply military-domain priors when a visually similar but non-combat vehicle appears in a military setting.

The observed qualitative errors can therefore be grouped into three categories: overly general captions that omit domain-relevant information, incomplete captions that identify the main scene but omit important objects or actions, and incorrect object classifications caused by visual similarity or target-domain bias. These examples also show why improvements in BLEU and ROUGE-L should be complemented by human evaluation of factual correctness. Since the qualitative analysis contains only a small number of selected images, it illustrates possible failure modes but does not establish their frequency in the broader target domain.

5. CONCLUSION

This paper extends a ResNet-152-LSTM image-captioning model trained on MS COCO to a specialized collection of military-exercise imagery. The principal contribution is an end-to-end domain-adaptation pipeline that combines automatic target-domain annotation with parameter-efficient fine-tuning. A pre-trained Florence-2 model generates five pseudo-reference captions per image, after which the original encoder-decoder checkpoint is adapted by updating the complete LSTM decoder and the encoder's projection and batch-normalization layers while keeping the convolutional ResNet backbone fixed.

Evaluation on military-domain images demonstrates a clear advantage over direct application of the source model. Fine-tuning improves BLEU-1, BLEU-2, BLEU-3, BLEU-4, and ROUGE-L by 18.87, 16.57, 10.61, 6.14, and 8.47 percentage points, respectively.

Beyond improving the quality of the generated captions, the results of this study also indicate the significant potential of such models for applications in modern defence technologies. During reconnaissance missions using unmanned platforms, video streams are typically transmitted continuously from the onboard camera to the ground control station. For example, transmitting a Full HD video stream (1920 × 1080 pixels, 30 fps) compressed using the H.264 codec at a bitrate of 5 Mb/s requires the transmission of approximately 625 kB of data per second. In contrast, a textual description of a single image generated by the proposed model occupies approximately 150 bytes. When processing 30 images per second, this corresponds to a data rate of only about 4.5 kB/s. These results open up the possibility of significantly reducing the amount of data transmitted over the communication link during the communication between the unmanned aerial vehicle and the ground control station, which may be particularly important under conditions of limited bandwidth.

Additional reductions in the required communication bandwidth can be achieved by applying an event-driven information transmission strategy. In this case, the system would continuously analyse images directly on the unmanned platform, while textual descriptions would be transmitted only if the generated caption contains predefined objects or activities of interest, such as military personnel, combat vehicles, helicopters, vessels, or other assets. This approach avoids transmitting a large number of irrelevant scenes, further reducing the communication load and improving the overall efficiency of the system.

Although this study focuses on automatic image captioning, the proposed approach also opens the possibility for the development of semantic communication systems, in which only information with operational significance is transmitted instead of the complete video stream. Such a concept could represent an important direction for the development of future intelligent reconnaissance systems, particularly in scenarios with limited communication bandwidth and the need for timely decision-making.

Several extensions should be considered in future work. First, the dataset should be expanded across a broader range of exercises, seasons, weather conditions, viewpoints, platforms, and equipment categories, with strict source-level partitioning to reduce correlations between training and test scenes. Second, a subset of automatically generated captions should be independently reviewed or rewritten by domain experts, and inter-annotator agreement should be reported. Finally, because defence-oriented captioning is safety-relevant, future work should quantify uncertainty, identify hallucinated objects or activities, and define a human-in-the-loop protocol under which generated descriptions support rather than replace operator judgment.

References

[1] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and Tell: A Neural Image Caption Generator," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3156-3164.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[4] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in Computer Vision - ECCV 2014, pp. 740-755, 2014.

[5] K. Xu et al., "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," in Proc. 32nd Int. Conf. Machine Learning (ICML), 2015, pp. 2048-2057.

[6] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[7] B. Xiao et al., "Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2024, pp. 4818-4829.

[8] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring Models and Data for Remote Sensing Image Caption Generation," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 4, pp. 2183-2195, 2018.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SWARM INTELLIGENCE: DIFFERENCES BETWEEN BIOLOGICAL PRINCIPLES AND ARTIFICIAL INTELLIGENCE CAPABILITIES

MILOŠ JOVANOVIĆ
Technical Test Center, Belgrade, milosjovanovic1967@gmail.com

JELENA VASILJEVIĆ
School of computing, Union university, Belgrade, jvasiljevic@raf.rs

IGOR MEDENICA
School of computing, Union university, Belgrade, medenicaigor@gmail.com

DRAGAN LAZIĆ
Technical Test Center, Belgrade, Belgrade, astazulazic@gmail.com

CALIN-OCTAVIAN MICLOSHINA
Faculty of Engineering, Resita, calin.miclosina@ubbcluj.ro

Abstract: *Swarm intelligence describes collective problem solving that emerges from interactions among multiple agents. This paper compares biological and artificial swarm principles and evaluates a stylized close-range defense scenario using two explicitly parameterized stochastic models. For each model, 10,000 Monte Carlo runs were performed with ten defenders and one intruder under identical initial conditions. The biological model represents fast local detection and recruitment, whereas the artificial model includes sensor-processing and communication delays. Under the stated assumptions, the biological swarm achieved a mean elimination time of 5.17 s versus 12.17 s for the artificial swarm and a lower mean defender-loss rate (1.13% versus 4.76%). These values are scenario-specific simulation outputs rather than physical measurements. The comparison illustrates the trade-off between rapid decentralized reaction and richer digitally mediated coordination.*

Keywords: *swarm intelligence, biological swarm, artificial swarm, multi-agent systems, Monte Carlo simulation.*

1. INTRODUCTION

Swarm intelligence refers to collective behavior that emerges from interactions among multiple relatively simple units without requiring continuous centralized control. In biological systems, such behavior is visible in social insects, fish schools, and other animal collectives, where local rules can produce robust group-level organization [1], [8], [12]. In this paper, an agent denotes an autonomous decision-making unit that perceives information from its environment or peers, selects actions, and contributes to a shared task [10], [11].

Bio-inspired computational methods such as Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) translated these principles into engineering and optimization algorithms [1]-[3]. Contemporary swarm robotics extends this idea to groups of physical or simulated robots that coordinate through local interaction, communication, and distributed decision-making [8], [9].

Biological and artificial swarms share principles of emergence, redundancy, and collective coordination, but they differ in communication bandwidth, memory, learning mechanisms, and access to global information. Artificial systems can additionally use machine learning, shared digital state, edge or cloud computation, and higher-level coordination mechanisms [10], [11]. These capabilities create new opportunities but also introduce latency, communication dependence, security risks, and possible single points of failure [13], [14].

The contribution of this paper is twofold. First, it provides a structured comparison of biological and artificial swarm principles using consistent terminology. Second, it evaluates a reproducible stochastic close-range defense model with explicitly stated assumptions, algorithms, and parameters. The simulation is intended as a comparative computational experiment rather than as a validated model of a particular biological species or robotic platform.

2. BIOLOGICAL PRINCIPLES OF SWARM INTELLIGENCE

Biological swarm intelligence emerges from local interactions among individuals and between individuals and the environment. No single individual normally maintains a complete representation of the global state; coordinated behavior instead arises through self-organization, feedback, and repeated local responses [1], [12].

Corresponding author: M. Jovanović;
E-mail: milosjovanovic1967@gmail.com
DOI: 10.5937/OTEH26067J

Decentralization is a defining property of many biological swarms. Individuals react to chemical, visual, acoustic, tactile, or motion-related cues. Ant foraging is a classical example of stigmergy, in which pheromone deposition modifies the environment and influences the subsequent behavior of other agents. The environment therefore acts as a distributed and continuously updated memory [1].

Redundancy and local interaction contribute to robustness and scalability. The loss of an individual agent usually does not terminate collective behavior, and coordination can continue without a dedicated central controller. Recent reviews of animal collective behavior and swarm robotics emphasize self-organization, resilience, adaptability, and locality as important principles transferred from biology to engineered swarms [8], [12].

These advantages coexist with limitations. Biological communication has restricted bandwidth, information propagation is local, and long-term adaptation is primarily evolutionary or behavioral rather than based on persistent digital memory. Accordingly, the biological model used later in this paper should be understood as a stylized model of rapid local collective response, not as a simulation of a specific animal species.

The distinction between biological principles and artificial implementation is therefore important: bio-inspired algorithms reproduce selected coordination mechanisms, whereas living swarms operate under biological sensing, communication, and energy constraints that cannot be transferred directly to engineered systems [8], [9].

3. ARTIFICIAL SWARM INTELLIGENCE AND ARTIFICIAL INTELLIGENCE-BASED COORDINATION

Artificial swarm intelligence refers to computational or robotic systems in which multiple autonomous agents cooperate to achieve a collective objective. Early approaches were strongly bio-inspired, including ACO and PSO, while modern swarm robotics combines distributed control, sensing, communication, and computation [2], [3], [8], [9].

Artificial agents can exchange digital messages directly and can access shared state through networked infrastructure. This allows faster dissemination of precise information than in many biological mechanisms and can provide partial or full knowledge of the system state. However, the achievable benefit depends on communication latency, bandwidth, synchronization, and fault tolerance [9], [14].

Machine learning further extends artificial swarm capabilities. Reinforcement learning and multi-agent reinforcement learning (MARL) allow agents to adapt policies from interaction data, while centralized training, decentralized execution, learned communication, and shared experience can support increasingly complex cooperative behavior [10], [11].

Higher-level coordination can also be introduced without removing local autonomy. An edge, cloud, or command component may aggregate observations, estimate global conditions, optimize task allocation, and periodically update agent policies. In a hybrid architecture, local agents still execute immediate actions while the higher-level component supports system-wide learning and planning [10], [14].

Persistent digital memory is another major difference from biological stigmergy. Artificial systems can store historical observations, compare alternative strategies, use predictive models, and redistribute learned policies. These functions are especially relevant to distributed sensing, vehicle coordination, multi-robot task allocation, and other applications that require information beyond the immediate local neighborhood [9], [11].

Artificial swarms may also support dynamic role assignment. Individual agents can be assigned to sensing, communication, pursuit, transport, or support roles according to mission state and available resources. Such explicit task decomposition can improve efficiency but increases algorithmic and communication complexity [11].

The same infrastructure that enables global information and learning can introduce vulnerabilities. Communication outages, compromised agents, malicious messages, or failure of shared infrastructure can disrupt collective behavior. Security of swarm communication and resilience to faults are therefore essential design considerations [13], [14].

For consistency, the term artificial swarm is used throughout the remainder of this paper. It covers decentralized, centralized-support, and hybrid implementations and avoids introducing a second label for the same class of engineered multi-agent systems.

4. KEY DIFFERENCES: BIOLOGICAL VS ARTIFICIAL SWARMS

Although artificial swarm intelligence is inspired by natural collective behavior, the two classes of systems should not be treated as equivalent. The most important differences concern the origin of coordination, communication mechanisms, memory, learning, system-level knowledge, and dependence on infrastructure [8], [9], [12].

Control and coordination provide the first distinction. Biological swarms are predominantly decentralized and rely on local interaction. Artificial swarms can also be decentralized, but they may additionally employ centralized-support or hybrid architectures in which global information is used for planning or learning while execution remains distributed [10], [14].

Communication is similarly different. Biological systems rely on local signals such as pheromones, movement, sound, or direct contact, whereas artificial agents may exchange structured digital messages at high speed. Digital communication can improve synchronization and task allocation, but it also creates dependence on networks and communication security [13].

Learning and memory provide another distinction. Biological adaptation is constrained by physiology,

behavior, and evolution, while artificial agents may use persistent databases, online learning, predictive models, and policy updates. These capabilities can increase flexibility but also create non-stationarity and coordination challenges in multi-agent learning [10], [11].

From a reliability perspective, the comparison is therefore a trade-off rather than a general ranking. Biological swarms emphasize locality, redundancy, and graceful degradation; artificial swarms can achieve richer planning and information sharing but must explicitly manage communication failures, cyber-security threats, and architectural bottlenecks [13], [14].

Table 1. Key Characteristics of Biological and Artificial Swarm Intelligence

Feature	Biological swarm	Artificial swarm
Control structure	Fully decentralized	Decentralized or hybrid (centralized support possible)
Agent complexity	Very limited individual intelligence	Programmable agents with learning capabilities
Communication	Indirect, low bandwidth (pheromones, movement, sound)	Direct, high-speed digital communication
Global knowledge	None	Partial or full global state awareness possible
Adaptation	Evolutionary, slow	Real-time learning and adaptation
Robustness	Very high	Moderate to high (architecture-dependent)
Scalability	Natural and implicit	Requires careful system design
Memory	Environmental (stigmergy-based)	Persistent digital memory
Optimization capability	Emergent and approximate	Explicit and goal-oriented

Table 1 summarizes these differences. The characteristics shown are general design tendencies rather than universal properties; a particular artificial swarm may deliberately reproduce fully decentralized biological principles or combine them with centralized support.

5. SIMULATION METHODOLOGY AND SWARM MODELS

The study uses two custom discrete-time stochastic multi-agent swarm models under the same close-range defense task: a biological-response model and an artificial-response model. One hostile intruder enters a defended local zone. The objective is to eliminate the intruder while recording response time and defender losses. The model is deliberately non-spatial: continuous coordinates, trajectories, and geometric collision dynamics are not simulated. Spatial effects are represented indirectly through detection, mobilization, and recruitment delays.

Each swarm contains $N = 10$ autonomous defender agents; the hostile intruder is an external entity and is not counted as a swarm agent. Each model was evaluated in 10,000 independent Monte Carlo runs (20,000 runs in total). Every run started with one intruder and the ten defender agents, used a discrete time step $\Delta t = 0.05$ s, and terminated when the intruder was eliminated or when the 120 s maximum duration was reached. A master random seed of 20260702 was used for reproducibility.

The simulation was implemented in Python using NumPy random-number generators. At the start of each run, detection time and mobilization delay were sampled from gamma distributions. When their combined delay elapsed, an initial subset of defenders became active. Each remaining inactive defender was then recruited independently in every time step with probability $P_R = 1 - \exp(-\rho\Delta t)$, where ρ is the model-specific recruitment rate.

Once defenders were active, the same event logic was used for both models. First, additional defenders could be recruited; second, intruder elimination was sampled using the aggregate elimination probability in (1); and third, if the intruder remained active, defender losses were sampled from a binomial distribution using the per-defender loss probability in (2). The number n of active defenders was updated after every recruitment or loss event.

Both swarms are represented by the same custom discrete-time stochastic multi-agent model evaluated by Monte Carlo simulation; they are not implementations of ACO or PSO. The biological-response model contains $N = 10$ defender agents and represents faster local detection, shorter mobilization, faster recruitment, and stronger local coordination. The artificial-response model likewise contains $N = 10$ defender agents but represents sensor-processing, communication, threat-classification, and task-allocation delays. Thus, the simulation algorithm and the number of swarm agents are identical, while the parameter sets differ as reported in Table 2.

For each run, the recorded outputs were intruder-elimination time, number of defenders lost, defender-loss percentage, and elimination success or failure. Across all runs, the mean, median, standard deviation, 90th and 95th percentiles of elimination time, mean defender loss, probability of losing at least one defender, and success rate were calculated.

The probability that the active defenders eliminate the intruder during an individual simulation interval was calculated as

$$P_E = 1 - e^{-\lambda c n \Delta t}, \quad (1)$$

where P_E is the probability of eliminating the intruder during the observed interval, λ is the basic elimination effectiveness of an individual defender, c is the coordination factor, n is the current number of active defenders, and Δt is the simulation time step.

The probability of an individual active defender being lost during a simulation interval was calculated as

$$P_L = 1 - e^{-\mu \Delta t}, \quad (2)$$

where P_L is the probability of losing an individual active defender during the interval and μ is the defender-loss rate.

Table 2 lists all parameter values used in the two models. Gamma distributions are parameterized by shape k and scale θ; their expected delay is $k\theta$. Parameters were selected as explicit scenario assumptions for a controlled comparison and were not fitted to measurements from a particular biological colony or physical robotic swarm.

Table 2. Parameters of the stochastic simulation models.

Parameter	Biological	Artificial
Defenders (N)	10	10
Time step Δt (s)	0.05	0.05
Max. time T_{max} (s)	120	120
Detection $\Gamma(k, \theta)$	(4, 0.38)	(4, 0.85)
Mobilization $\Gamma(k, \theta)$	(3, 0.15)	(3, 0.40)
Initial active (n_0)	3	2
Recruitment ρ (s^{-1})	1.80	0.80
Elimination λ (s^{-1})	0.027	0.016
Coordination (c)	1.30	1.00
Loss μ (s^{-1})	0.0038	0.0075

The biological-response model starts with $n_0 = 3$ active defenders after detection and mobilization, recruits nearby defenders at a higher rate ρ, and uses a coordination factor $c > 1$ to represent efficient local collective action. The artificial-response model starts with $n_0 = 2$ active defenders and uses longer sensing and mobilization delays, slower recruitment, and a lower effective elimination rate λ.

The simulation does not encode a specific ant, bee, hornet, drone, or robot platform. Instead, it isolates the consequences of two contrasting coordination assumptions under a single local-defense task. Consequently, the numerical results must be interpreted only within the stated model and parameter ranges.

This explicit parameterization is important because the same qualitative labels, such as faster detection or stronger coordination, can produce different outcomes when assigned different numerical values. Reproducibility therefore depends on reporting both the stochastic algorithm and the complete parameter set.

The comparative design intentionally holds the initial threat, defender population, time step, stopping rule, and output metrics constant. This allows observed differences to be attributed to the stated model parameters rather than to different initial conditions.

6. RESULTS AND DISCUSSION

Both models eliminated the intruder in all 10,000 runs. The biological model achieved a mean elimination time of 5.17 s, compared with 12.17 s for the artificial model. This corresponds to a 57.5% lower mean elimination time under the stated parameterization.

The median elimination times were 4.40 s and 10.30 s, while the standard deviations were 3.11 s and 7.40 s for the biological and artificial models, respectively. Figure 1 presents the cumulative distribution of elimination time across all simulation runs.

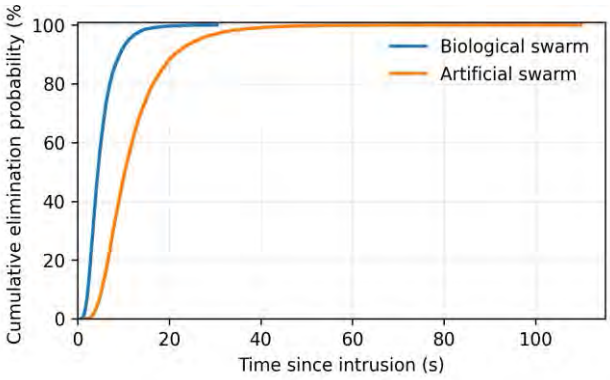


Figure 1. Cumulative probability of intruder elimination for the biological and artificial swarm models.

The biological curve rises earlier because the model combines shorter detection and mobilization delays with faster recruitment. Ninety percent of biological-model runs completed elimination within 9.20 s, whereas the artificial model required 21.20 s. At the 95th percentile, the corresponding values were 11.20 s and 26.45 s.

The wider artificial-model distribution reflects the additional variability introduced by longer sensing, processing, communication, and mobilization delays. This finding should not be interpreted as a universal property of artificial swarms; it follows directly from the close-range scenario and the parameter values reported in Table 2.

Figure 2 compares the mean elimination times with one-standard-deviation error bars. The mean difference was 7.00 s. In this scenario, the intruder therefore remained active for more than twice as long on average in the artificial model.

The result is consistent with the model design: immediate local information is sufficient for the task, so the biological parameterization gains little from global state while avoiding the additional stages assigned to artificial coordination. This is a task-specific advantage of the modeled response chain rather than evidence that biological swarms are generally more capable.

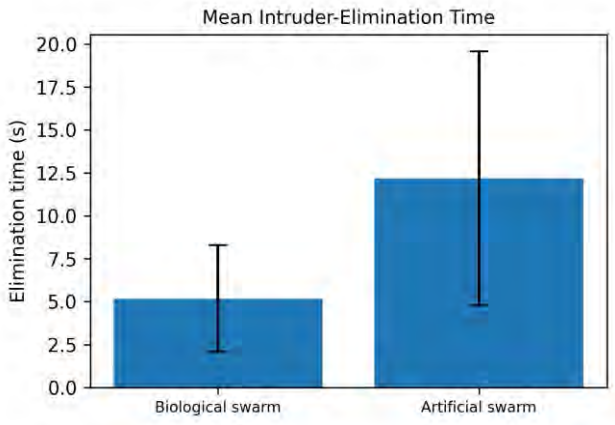


Figure 2. Mean intruder-elimination time for the biological and artificial swarm models; error bars show one standard deviation.

Defender losses show the same scenario-specific tendency. The biological model lost an average of 0.113 defenders per intrusion, or 1.131% of the initial population. The artificial model lost 0.476 defenders, or 4.759%. Figure 3 summarizes the mean loss rates.

The relative reduction in mean defender losses was 76.2%. The probability of losing at least one defender was 9.81% for the biological model and 31.99% for the artificial model, making such a loss approximately 3.26 times more frequent in the artificial-model runs.

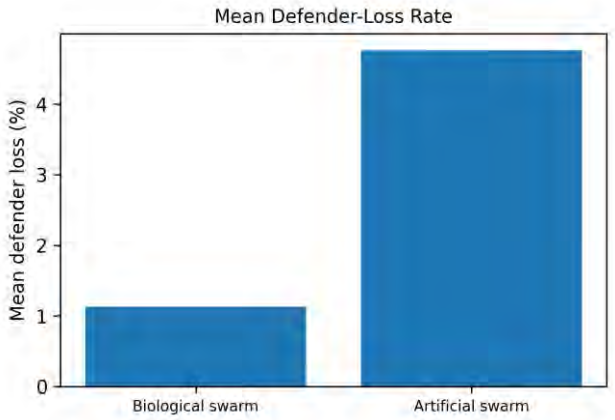


Figure 3. Mean defender-loss rate for a nest initially containing ten defenders.

These simulated losses arise because longer engagements expose active defenders to the per-interval loss process for more time. The loss difference is therefore linked to both the elimination dynamics and the model-specific loss-rate parameter; it is not an independently measured property of biological or robotic systems.

The simulation directly supports only conclusions about the modeled close-range defense task: under the parameters in Table 2, faster local detection, recruitment, and coordination produce shorter engagements and fewer losses. It does not test tasks that require global mapping, long-distance coordination, prediction, persistent memory, or dynamic allocation across multiple geographically separated threats.

Advantages commonly associated with artificial swarms in those broader tasks are therefore literature-based design capabilities rather than outputs of the present simulation. Recent work on MARL and multi-robot systems shows how learned policies, shared information, and centralized-training/decentralized-execution architectures can support complex coordination [10], [11], while resilience research emphasizes the need to preserve local autonomy when communication or shared infrastructure fails [14].

The main limitation of the present experiment is the use of synthetic scenario parameters rather than empirical calibration. The model is also non-spatial and does not simulate radio propagation, packet loss, physical trajectories, heterogeneous agents, or multiple simultaneous intruders. These simplifications make the experiment reproducible and interpretable but limit external validity.

Accordingly, the results should be interpreted as a controlled comparison of two coordination assumptions. They demonstrate how local-response and communication-mediated delays can affect a stochastic defense model, not that one class of swarm is universally superior to the other.

For immediate local defense, decentralized reaction can be advantageous when local information is sufficient and communication infrastructure is unnecessary. In contrast, artificial swarms may be preferable when the task depends on persistent memory, remote information, adaptive task allocation, or coordinated action over larger spatial scales [9]-[11].

This distinction reconciles the simulation results with the broader comparison in Table 1: the present experiment evaluates one operating regime, while the general strengths and weaknesses of biological and artificial swarms depend on mission requirements and system architecture.

7. CONCLUSION AND FUTURE RESEARCH

This paper compared biological and artificial swarm intelligence at both conceptual and simulation levels. Biological swarms emphasize local interaction, self-organization, and redundancy, whereas artificial swarms can add digital communication, persistent memory, machine learning, and higher-level coordination. These additional capabilities expand the range of possible tasks but introduce communication, security, and architectural dependencies.

In the explicitly parameterized close-range defense model, the biological swarm achieved a mean elimination time of 5.17 s compared with 12.17 s for the artificial swarm, while the mean defender-loss rates were 1.13% and 4.76%, respectively. Because the parameters are synthetic and the model is non-spatial, these results are evidence about the stated computational scenario only.

The central conclusion is therefore conditional: rapid decentralized reaction is advantageous when a threat is local and immediate, whereas richer artificial coordination can become advantageous when tasks

require information aggregation, learning, prediction, or system-wide resource allocation. Neither architecture is universally superior.

Future work should calibrate the model against empirical observations or physical multi-robot experiments, introduce explicit spatial motion and communication-network models, and evaluate multiple simultaneous threats. Additional experiments should compare decentralized and hybrid artificial architectures under packet loss, latency, agent failure, and adversarial communication conditions [13], [14].

A further research direction is to replace fixed artificial-swarm parameters with learned policies based on MARL and to test whether adaptive task allocation can compensate for communication and processing delays in close-range scenarios [10], [11]. Such extensions would provide a stronger basis for determining when biological principles should be preserved directly and when artificial-intelligence-based coordination provides measurable benefit.

References

[1] E. Bonabeau, M. Dorigo, and G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press, 1999.

[2] M. Dorigo, V. Maniezzo, and A. Colorni, “Ant system: Optimization by a colony of cooperating agents,” IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 26, no. 1, pp. 29-41, 1996.

[3] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proceedings of the IEEE International Conference on Neural Networks, 1995, pp. 1942-1948.

[4] E. Şahin, “Swarm robotics: From sources of inspiration to domains of application,” in Swarm Robotics, LNCS 3342, Springer, 2005, pp. 10-20.

[5] L. Busoniu, R. Babuska, and B. De Schutter, “Multi-agent reinforcement learning: An overview,” in Innovations in Multi-Agent Systems and Applications - 1, Springer, 2010, pp. 183-221.

[6] A. Brambilla et al., “Swarm robotics: A review from the swarm engineering perspective,” Swarm Intelligence, vol. 7, pp. 1-41, 2013.

[7] M. Jovanovic, J. Vasiljevic, D. Lazic, I. Medenica, and D. Nagamalai, “Swarm Robotics in the Military Operations: Challenges and Opportunities,” Computer Science & Information Technology, 2025, doi: 10.5121/CSIT.2025.150519.

[8] P. G. F. Dias, M. C. Silva, G. P. Rocha Filho, P. A. Vargas, L. P. Cota, and G. Pessin, “Swarm Robotics: A Perspective on the Latest Reviewed Concepts and Applications,” Sensors, vol. 21, no. 6, 2062, 2021, doi: 10.3390/s21062062.

[9] C. Calderón-Arce, J. C. Brenes-Torres, and R. Solis-Ortega, “Swarm Robotics: Simulators, Platforms and Applications Review,” Computation, vol. 10, no. 6, 80, 2022, doi: 10.3390/computation10060080.

[10] S. Gronauer and K. Diepold, “Multi-agent deep reinforcement learning: a survey,” Artificial Intelligence Review, vol. 55, pp. 895-943, 2022, doi: 10.1007/s10462-021-09996-w.

[11] J. Orr and A. Dutta, “Multi-Agent Deep Reinforcement Learning for Multi-Robot Applications: A Survey,” Sensors, vol. 23, no. 7, 3625, 2023, doi: 10.3390/s23073625.

[12] H. Duan, M. Huo, and Y. Fan, “From animal collective behaviors to swarm robotic cooperation,” National Science Review, vol. 10, no. 5, nwad040, 2023, doi: 10.1093/nsr/nwad040.

[13] L. Chen and S.-L. Ng, “Securing emergent behaviour in swarm robotics,” Journal of Information Security and Applications, vol. 64, 103047, 2022, doi: 10.1016/j.jisa.2021.103047.

[14] D. M. Bossens, S. Ramchurn, and D. Tarapore, “Resilient Robot Teams: a Review Integrating Decentralised Control, Change-Detection, and Learning,” Current Robotics Reports, vol. 3, pp. 85-95, 2022, doi: 10.1007/s43154-022-00079-4.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

APPLICATION OF THE BARTLETT METHOD FOR TARGETS ANGLE ESTIMATION IN FMCW RADAR SYSTEMS

BORKO ĐAKOVIĆ
Applied Mathematics and Electronics Center, Military Academy, Belgrade,
borko.djakovic@vs.rs, ORCID: 0009-0000-4662-3981

SLOBODAN SIMIĆ
Military Academy, Belgrade,
slobodan.simic@va.mod.gov.rs, ORCID: 0000-0002-6846-4886

NIKOLA DANILOVIĆ
Military Technical Institute, Belgrade,
danilovicn5@gmail.com, ORCID: 0009-0007-1345-7527

Abstract: *This paper presents the application of the Bartlett method for estimating the angle of arrival of targets using an FMCW radar system. After organizing the samples and forming the radar data cube, a 2D FFT algorithm is first applied to obtain the Range–Doppler matrix. Subsequently, an IFFT is performed on the samples corresponding to the same target range, and the Bartlett method is then applied to estimate the angle of arrival. The algorithm is repeated for all range resolution cells. Range–Angle map obtained using both the Angle-DFT method and the Bartlett method are presented, and their results are comparatively analyzed. The simulation is conducted for multiple targets with different parameters and varying SNR levels, while experimental validation is performed using an FMCW TDM MIMO radar system.*

Keywords: *Bartlett method, Angle estimation, FMCW TDM MIMO Radar, Angle DFT, Range-Angle map.*

1. INTRODUCTION

The need for accurate estimation of fundamental target parameters, such as range, velocity, and angle of arrival, has led to the widespread adoption of Frequency-Modulated Continuous-Wave (FMCW) radar systems in modern applications. FMCW radars are used in the mode of precise target tracking, especially in military applications such as guidance radars for missile control and ballistic radars [1]. Also, this type of radar has found extensive use in medical applications, including biosignal monitoring [2] and tumor detection [3], in the detection of unmanned aerial vehicles (UAVs) with a low radar cross-section [4], and, most notably, in the automotive industry, particularly in autonomous driving systems [5]. Owing to their high time-bandwidth (tB) product, FMCW radars provide highly accurate and precise range and velocity estimation, whereas angle estimation and angular resolution are directly determined by the antenna aperture and the number of receiving antennas [6].

In practical scenarios, however, the implementation of a large number of receiving antennas is often impractical due to space constraints, cost considerations, and the need to optimize hardware resources. When conventional angle estimation techniques, such as the Angle Discrete Fourier Transform (Angle DFT), are employed, even Multiple-Input Multiple-Output (MIMO) radar systems, which utilize multiple transmitting antennas and virtual array formation techniques, fail to achieve satisfactory angle estimation performance.

Therefore, the use of more advanced angle-of-arrival (AoA) estimation algorithms is essential. At the same time, there is a strong demand for computationally efficient algorithms that do not require excessive processing power, thereby enabling their practical implementation in real-time radar systems.

The Bartlett method is a conventional beamforming technique for estimating the direction of arrival (DoA) of signals impinging on an antenna array, and it is most widely used in direction-finding applications. Moreover, the method can also be applied when two antenna arrays need to be precisely aligned toward each other, ensuring accurate orientation and minimizing signal attenuation. However, with appropriate modifications, it can also be applied to radar systems. The method is based on beamforming, electronically steering the array toward a desired direction, and calculating the received signal power. Its main advantages include straightforward implementation, low computational complexity, fast execution, and robust performance in the presence of noise.

The paper is organized as follows. The first section introduces the research problem, discusses the motivation for the study, and provides a brief overview of the Bartlett method for target angle-of-arrival estimation. The second section presents the signal and system models used throughout the study. The third section describes the

Corresponding author: B. Đaković;
E-mail: borko.djakovic@vs.rs
DOI: 10.5937/OTEH26068D

mathematical formulations of the Angle Discrete Fourier Transform (Angle DFT) algorithm and the Bartlett method for target angle estimation in radar systems. The fourth section presents simulation results evaluating the performance of the proposed algorithms in terms of angle estimation accuracy for multiple targets, the ability to resolve closely spaced targets in the angular domain, and robustness to noise. The fifth section provides an experimental validation of the algorithms using an FMCW MIMO radar platform, followed by the conclusions of the study..

2. RADAR SYSTEMS AND SIGNAL MODEL

2.1. Radar System Model

Fig. 1 illustrates the architecture of the Time Division Multiplexing (TDM) Multiple-Input Multiple-Output (MIMO) Frequency-Modulated Continuous-Wave (FMCW) radar system employed in this study..

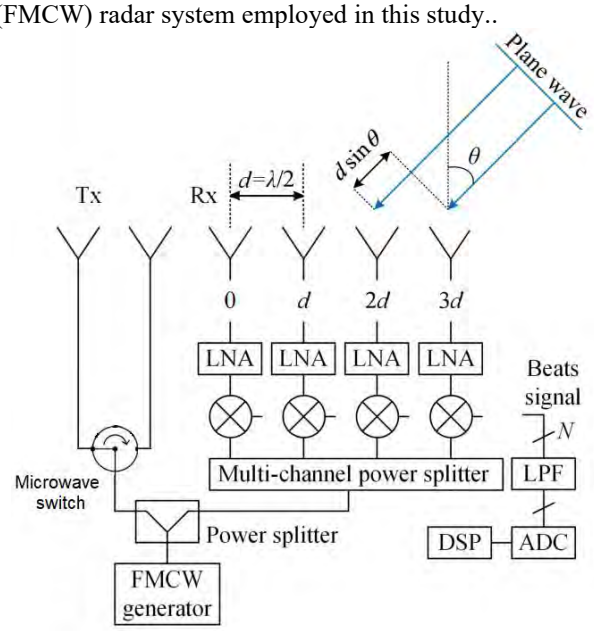


Figure 1. Radar System Model

The radar system consists of two transmitting antennas and four receiving antennas. The FMCW signal is routed to the transmitting antennas through a microwave switch, which controls the transmission sequence in the time domain. Specifically, the signal is transmitted alternately from the first and the second transmitting antenna. After being reflected by the target, the echo signal is received by a linear antenna array. Since the target is assumed to be located in the far field, the reflected wavefront can be approximated as a plane wave at the receiving array.

The received signal is first amplified by a low-noise amplifier (LNA) and then fed into an I/Q demodulator. Subsequently, it passes through a low-pass filter, producing the intermediate-frequency (IF) signal. After analog-to-digital conversion, the I/Q samples are processed digitally. The acquired samples are arranged into a three-dimensional data structure, commonly referred to as the radar data cube. One dimension contains the samples acquired within a single chirp, the second dimension contains samples from multiple chirps, and the third dimension contains samples collected by multiple receiving antennas.

Two radar data cubes are formed, each corresponding to one transmitting antenna. By concatenating the received signals, a virtual antenna array is created, exhibiting the characteristics of a system with a single transmitting antenna and eight receiving antennas. This enables the construction of a unified radar data cube. Since the chirp duration is very short, typically on the order of a few milliseconds, the target can be assumed to remain within the same range resolution cell during the acquisition interval, making Doppler phase compensation unnecessary.

The target range is estimated from the beat frequency of the IF signal, which depends on the target position, whereas the target angle is determined from the phase delay of the received signal across the antenna array [7].

2.2. Radar Signal Model

To construct the Range–Angle map at a given time instant, the signal samples corresponding to a single chirp and multiple receiving antennas are considered. These signal samples can be analytically represented as a two-dimensional complex exponential signal [8]:

$$Y(m,n) = \sum_{l=1}^{L} \alpha_l e^{j\varphi_l} e^{j2\pi \frac{2R_l B}{cT_c} m T_S} e^{j\frac{2\pi}{\lambda} d n \sin\theta_l} + n_n(m,n), \quad (1)$$

where $\alpha_l e^{j\varphi_l}$ denote the complex amplitude and phase, R_l range, Θ_l angle of the l-th target out of a total L targets, B bandwith of FMCW signal, c the propagation velocity of the electromagnetic (EM) wave, which can be assumed to be equal to the speed of light in air, Tc chirp duration, T_S sampling period, λ wavelength of the transmitted signal, inter-element spacing between the receiving antennas $d = \lambda/2$, $n_n(n,m)$ noise , $n = 0...N_{R_x}-1$, where N_{R_x} is number of receiving antennas, $m = 0... N_o-1$, where N_o is number of samples in a chirp, which is given by $N_o = T_c/T_S$. The sample matrix Y has dimensions $[N_o \, x \, N_{R_x}]$.

3. ANGLE ESTIMATION METHODS

3.1. Angle DFT Algorithm

The Angle DFT algorithm is typically applied after the Range DFT processing stage, once the beat frequency and target range have been estimated using the Discrete Fourier Transform (DFT). The Angle DFT algorithm is based on the application of the Discrete Fourier Transform to estimate the phase shift between signals received at different antenna elements. This phase shift arises due to the time delay associated with the additional propagation path length when the target is located at a non-zero angle of arrival.

The angle of the target is given by [7]:

$$\Theta_u = sin^{-1}(u\frac{\lambda}{Nd}), \quad (2)$$

where is $-\frac{N}{2} \leq u < \frac{N}{2}$, where N denotes the number of DFT coefficients. The angular measurement range and angular resolution are given by:

$$-sin^{-1}(\frac{\lambda}{2d}) \leq \theta < sin^{-1}(\frac{\lambda}{2d}), \quad (3)$$

$$\theta_{\text{res}} = 2 \cdot sin^{-1}(\frac{\lambda}{2N_{R_x} d\, cos(\tilde{\theta})}), \quad (4)$$

where $\tilde{\theta}$ is the central angle under which the target is observed. The angular resolution does not depend on the number of DFT coefficients, but solely on the number of receiving antennas.

3.2. Bartlett Method

If the Bartlett method were applied directly to the radar data cube, the resulting output would consist of the angles of arrival corresponding to all targets, and the estimated angles could not be associated with specific targets. Therefore, before applying the Bartlett method, it is necessary to perform target selection in the range domain. Subsequently, the method is applied to all range resolution cells corresponding to the same range bin.

Let Z be a three-dimensional radar data cube of dimensions [m x p x n], such that:

$$Z(:,:,k) = Y_k,\ k = 1, 2, \ldots, p, \quad (5)$$

where Y_k represents the matrix Y corresponding to the k - th chirp within a frame. In order to perform target selection in the range domain, it is first necessary to apply the Range DFT and Doppler DFT algorithms [9] across all antennas, yielding the range–Doppler matrices RD_n of dimensions [m x p]:

$$RD_n = fft2\ (Z\ (:, :, n), \quad (6)$$

where is $n = 0...N_{R_x}-1$, and *fft2* represents the two-dimensional Fast Fourier Transform (2D FFT) algorithm. In order to obtain signal samples corresponding to a single range resolution cell m, it is necessary to apply the inverse Fourier transform only to the selected samples. Accordingly, a vector $R_{n,m}$, of dimensions [1 x p] is obtained, which is given by:

$$R_{n,m} = ifft(\ RD_n(m,:)), \quad (7)$$

where *ifft* represents the inverse Fast Fourier Transform algorithm. When the matrices $R_{n,m}$ corresponding to a single range resolution cell are computed for all antennas in the antenna array, a matrix x_m of dimensions [n x p], can be obtained, which is given by:

$$x_m = [R_{1,m}; R_{2,m}; \ldots; R_{n,m}]^T. \quad (8)$$

The Bartlett method is based on the principle of maximizing the average output power of a conventional delay-and-sum beamformer as a function of the signal's direction of arrival. Accordingly, the Bartlett criterion function for target angle estimation within a single range resolution cell is given by [10]:

$$P_m\ (\theta,\varphi) = v^H\ (\theta,\varphi) X_m\ v\ (\theta,\varphi), \quad (9)$$

where X_m the estimated covariance matrix of the antenna array signal for a single range resolution cell is given by:

$$X_m = \frac{1}{p} x_m x_m^H, \quad (10)$$

where x_m^H is the Hermitian transpose of the matrix x_m. The vector $v(\theta,\varphi)$ is the steering vector corresponding to a given direction and is given by::

$$v(\theta,\varphi) = e^{-j2\pi\ a\ P/\lambda_0}, \quad (11)$$

where a represents the vector that determines the target direction of arrival:

$$a = -[sin(\varphi)cos(\theta),\ cos(\theta)cos(\varphi),\ sin(\theta)], \quad (12)$$

the vector P has dimensions [3 x N_{R_x}] and defines the positions of the receiving antennas in a Cartesian coordinate:

$$P = \begin{bmatrix} x_1 & \cdots & x_{N_{Rx}} \\ y_1 & \cdots & y_{N_{Rx}} \\ z_1 & \cdots & z_{N_{Rx}} \end{bmatrix}, \quad (13)$$

and λ_0 wavelength at the radar operating frequency. Vector $v^H(\theta,\varphi)$, denotes the Hermitian transpose of the steering vector $v(\theta,\varphi)$.

The Bartlett criterion function is evaluated over the set of search angles, and after applying it to all range resolution cells, a unified Range–Angle map is obtained.

4. SIMULATION

For simulation purposes, a radar data cube was generated according to Equation (1) and (5) of the radar signal model. The radar parameters used in both simulation and experimental validation are as follows: operating frequency f_c = 24 GHz; signal bandwidth B =1 GHz; chirp duration T_c = 0,5 ms; sampling frequency f_s = 64 KHz; number of samples per chirp N_o = 32, and number of antennas in the virtual antenna array N_{R_x} = 8. The number of chirps per frame is Nc = 128. These parameters provide an unambiguous radar range of up to 4.8 m and enable angle estimation over the range of -90º do 90º.

The simulation results are presented in the form of Range–Angle maps obtained using both methods. For the Bartlett method, a range step of 0.15 m and an angular step of 0,5º were used, resulting in a Range–Angle map of dimensions [32 x 361]. For a fair comparison, the DFT coefficients in the Angle DFT method were computed using the same number of points.

Table 1 presents the target parameters used in the simulation.. The parameter A denotes the normalized amplitude of the received target echo signal, which can be associated with the radar cross section (RCS) of the target. The parameter r represents the target range, while v denotes the radial velocity of the target, which can take both positive and negative values depending on whether

the target is moving toward or away from the radar. The parameter Θ represents the target angle, which can also be positive or negative depending on whether the target is located to the left or right of the radar boresight.

In these applications, target velocity is not of primary importance, as Range–Doppler maps are not considered. The targets are generated with different amplitudes. Furthermore, targets 2 and 3, as well as 4 and 5, are arranged in pairs such that the angular separation within each pair is identical, while the central observation angles differ, in order to investigate the effect on angular resolution performance. The targets are generated without the presence of noise.

Table 1. Target Parameters

R. br	*A* [V]	*r* [m]	*v* [m/s]	Θ [º]
1.	1	4	-5	50
2.	0.9	3	3	-20
3.	0.8	3	2	-40
4.	0.7	1	-2	45
5.	0.6	1	-1.5	65

Fig. 2 shows a three-dimensional representation of the Range–Angle matrix obtained using the Angle DFT method. The matrix is normalized with respect to its maximum value, and the target angles can be estimated by identifying local peaks. An estimation error in the angle is clearly observable, particularly for targets located at larger angles.

It can also be observed that targets 4 and 5 are not successfully resolved, even though their angular separation is the same as that of targets 2 and 3. This indicates that, as the central observation angle increases, the effective angular resolution cell also increases, which in turn degrades the ability to resolve closely spaced targets.

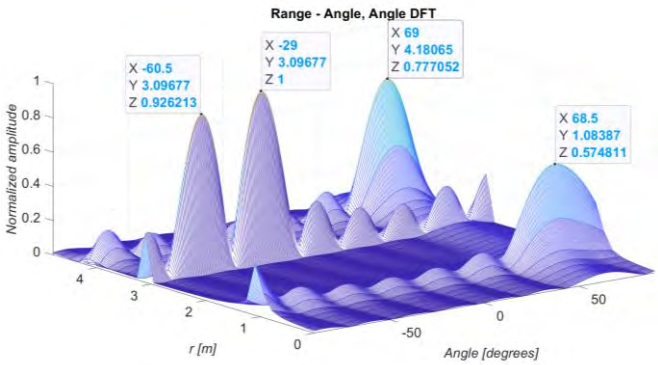


Figure 2. Range–Angle Matrix – Angle DFT Algorithm

Fig. 3 shows a three-dimensional representation of the Range–Angle matrix obtained using the Bartlett method. In contrast to the Angle DFT method, the Bartlett method exhibits significantly higher accuracy in target angle estimation. However, it can be observed that targets 4 and 5 are still not successfully resolved in this case as well, leading to the conclusion that the Bartlett method does not improve angular resolution.

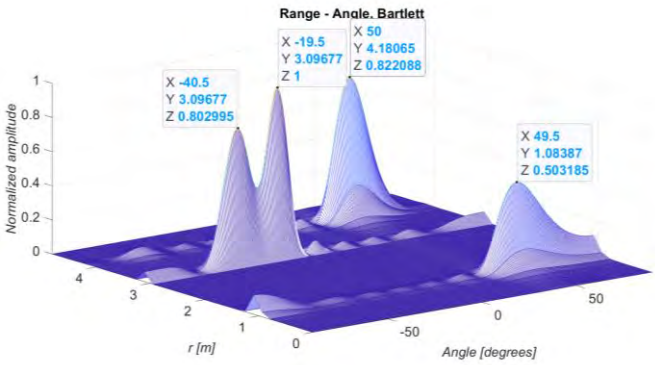


Figure 3. Range–Angle Matrix – Bartlett Method

Table 2 presents the absolute angle estimation errors obtained using the considered methods. It can be observed that the Bartlett method, compared to the Angle DFT algorithm, provides significantly more accurate target angle estimates. For targets that are successfully resolved, the estimation error is on the order of the angular step size used in the angle estimation procedure.

Table 2. Absolute angle estimation error

Cilj	1.	2.	3.	4.	5.
Angle DFT	19º	-9º	-20.5º	23.5º	3.5º
Bartlett	0º	0.5º	-0.5º	4.5º	-15.5º

FMCW radars enable signal detection even at low signal-to-noise ratios (SNR) [11]. Fig. 4 shows Range–Angle heatmaps for the targets listed in Table 1 under additive noise conditions applied to the radar data cube, with signal-to-noise ratio (SNR) levels of 10 dB, 0 dB, and −10 dB. The targets in the heatmaps are marked with red circles.

It can be concluded that the Bartlett method exhibits significantly higher robustness to noise, as targets can still be reliably detected even at low SNR values, whereas this is not the case for the Angle DFT algorithm.

.

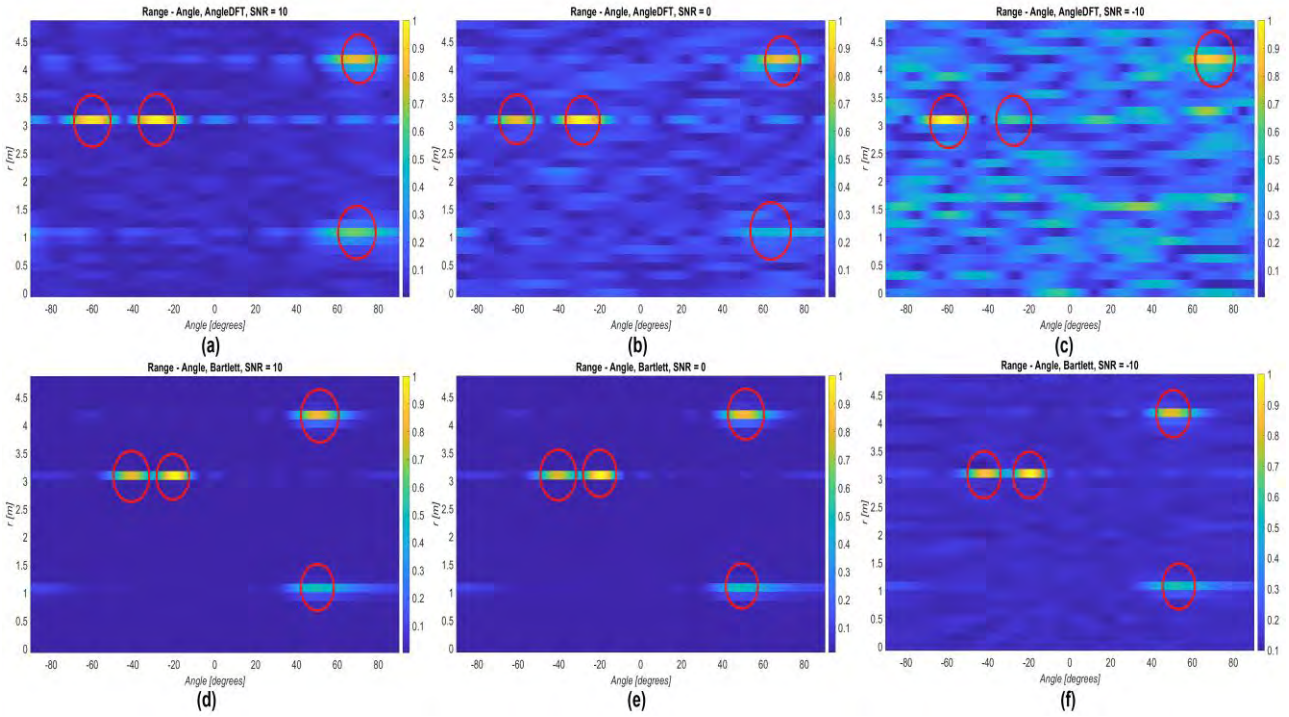


Figure 4. Range–Angle Maps (a) Angle DFT, SNR = 10 dB; (b) Angle DFT, SNR = 0 dB; (c) Angle DFT, SNR = −10 dB; (d) Bartlett method, SNR = 10 dB; (e) Bartlett method, SNR = 0 dB; (f) Bartlett method, SNR = −10 dB

5. EXPERIMENTAL VALIDATION

For the experimental validation, a MIMO radar platform PUP_DUAL24P_T2R4 manufactured by Luswave Technology was used (Fig. 5).

Figure 5. FMCW MIMO Radar platform

The platform operates in the K frequency band and consists of two transmitting and four receiving antennas. It generates an FMCW waveform and enables the adjustment of parameters such as chirp duration, number of samples per chirp, operating frequency, and signal bandwidth, as well as the capability to record I/Q samples for subsequent signal processing [12].

The radar target in the experiment was a person moving toward the radar from its right-hand side at an angle of approximately 30°. Fig. 6 shows the Range–Angle map obtained using the Angle DFT algorithm, while Fig. 7 presents the results obtained using the Bartlett method. The experimental results demonstrate that, in addition to providing more accurate angle estimation, the Bartlett method exhibits significantly higher robustness to noise.

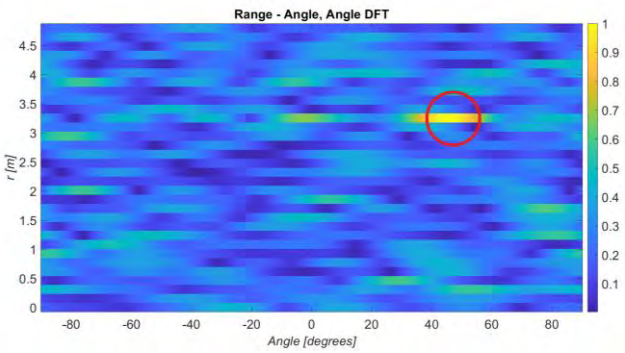


Figure 6. Range-Angle map, Experimental validation - Angle DFT

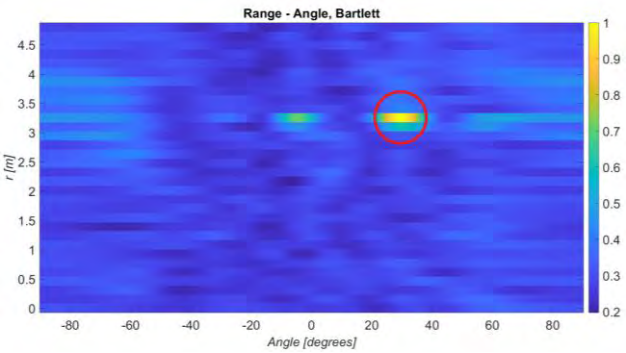


Figure 7. Range-Angle map, Experimental Validation – Bartlett Method

6. CONCLUSION

In this paper, the efficiency of the Bartlett method in target angle estimation and Range–Angle map generation using FMCW radar has been demonstrated through both simulation and experimental validation. It is confirmed that, compared to the Angle DFT algorithm, the Bartlett method introduces a negligible error in angle estimation, on the order of the angular search step size. It is also shown to be significantly more robust to noise and capable of producing satisfactory results at low signal-to-noise ratio (SNR) levels.

Furthermore, it is established that the use of the Bartlett method does not improve angular resolution, as it is fundamentally determined by the number of antennas in the antenna array. It is also observed that the separation of closely spaced targets in angle is somewhat more challenging, which requires tuning of the optimal detection threshold for angle estimation and opens possibilities for further research.

Overall, the Bartlett method represents a significantly more efficient approach for Range–Angle map generation compared to the Angle DFT algorithm. Its importance is further supported by its low computational complexity, as, in addition to FFT and IFFT operations, all other mathematical operations are trivial, making it suitable for implementation in compact systems.

Acknowledgment

In this paper, part of the results derived from the project CPME/P1/26-28, supported by the Ministry of Defence and the Serbian Armed Forces, have been used.

References

[1] A. A. Salama, M. A. Hussein, "Enhanced projectile path estimation using multi-vehicle FMCW radar sensors", *Scientific Reports,* vol. 16, Article 453, 2026.

[2] *M. Alizadeh,* G. Shaker, J. C. M. D. Almeida, P. P. Morita and S. Safavi-Naeini, "Remote Monitoring of Human Vital Signs Using mm-Wave FMCW Radar," in *IEEE Access*, vol. 7, pp. 54958-54968, 2019, doi: 10.1109/ACCESS.2019.2912956J.

[3] D. W. Bliss and K. W. Forsythe, "MIMO Radar Medical Imaging: Self-Interference Mitigation for Breast Tumor Detection," *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2006, pp. 1558-1562, doi: 10.1109/ACSSC.2006.355020.

[4] Á. D. de Quevedo, F. I. Urzaiz, J. G. Menoyo and A. A. López, "Remotely Piloted Aircraft Detection with Persistent Radar," *2018 15th European Radar Conference (EuRAD)*, Madrid, Spain, 2018, pp. 317-320, doi: 10.23919/EuRAD.2018.8546546.

[5] I. Bilik, O. Longman, S. Villeval and J. Tabrikian, "The Rise of Radar for Autonomous Vehicles: Signal Processing Solutions and Future Research Directions," in *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 20-31, Sept. 2019, doi: 10.1109/MSP.2019.2926573.

[6] Richards, Mark A. *Fundamentals of radar signal processing*. Vol. 1. New York: Mcgraw-hill, 2005.

[7] B. Đaković, S. Simić, L. Trivundža and A. Ristić, "Digital Signal Processing in MIMO Radars with Time Multiplexed Transmit Signals" in *Military Technical Courier,* vol. 2/2025, 2025, doi: 10.5937/vojtehg73-54972.

[8] G. O. Manokhin, Z. T. Erdyneev, A. A. Geltser and E. A. Monastyrev, "MUSIC-based algorithm for range-azimuth FMCW radar data processing without estimating number of targets," *2015 IEEE 15th Mediterranean Microwave Symposium (MMS)*, Lecce, Italy, 2015, pp. 1-4, doi: 10.1109/MMS.2015.7375471.

[9] X. Li, X. Wang, Q. Yang and S. Fu. (2021). Signal Processing for TDM MIMO FMCW Millimeter-Wave Radar Sensors, in IEEE Access, vol. 9, pp. 167959-167971, 2021, doi: 10.1109/ACCESS.2021.3137387.

[10] M. Erić, N. Vukmirović: *Uvod u obradu signala sa antenskih nizova,* Akademska misao, Beograd, 2019.

[11] S. Simić. FMCW radar signal analysis by time-frequency representations and correlation techniques. *Vojnotehnički glasnik*, 2007, 55.4: 400-413. doi: 10.5937/vojtehg0704400S

[12] Luswave Technology *PUP_DUAL24P_T2R4 Datasheet,* 2020. Dostupno na: https://luswave.com/PUP_EN24P_T2R4

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE ANALYSIS OF YOLO MODELS FOR OBJECT DETECTION IN IMAGES COMPRESSED WITH JPEG, BPG, AND VVC

NIKOLA BUKVIĆ
University of Defence in Belgrade, Military Academy, bukvicnikol@gmail.com

BOBAN BONDŽULIĆ
University of Defence in Belgrade, Military Academy, bondzulici@yahoo.com

DIMITRIJE BUJAKOVIĆ
University of Defence in Belgrade, Military Academy, dimitrijebujakovic@gmail.com

MILENKO ANDRIĆ
University of Defence in Belgrade, Military Academy, andricsmilenko@gmail.com

BOBAN PAVLOVIĆ
University of Defence in Belgrade, Military Academy, boban.pavlovic@va.mod.gov.rs

Abstract: *Object detection represents one of the key areas of visual data processing and has significant applications in the analysis of satellite and aerial imagery, intelligent transportation systems, surveillance, and autonomous systems. In modern applications, visual data are often compressed using lossy compression techniques to reduce storage and transmission requirements. However, compression leads to image quality degradation, which may negatively affect the performance of object detection algorithms. In this paper, a comparative analysis of the performance of YOLO models on images compressed using different methods and compression levels is conducted. The experimental analysis is performed on a set of 90 selected images from the DOTA v1.0 dataset. The images are compressed using JPEG, BPG, and VVC methods, with 11 representative compression factor values applied for each method. Object detection was carried out using YOLOv5 and YOLOv7–YOLOv12 models. Detection performance is evaluated using standard metrics, including precision, recall,* mAP_{50}*–*mAP_{95}*, and overall mAP. Particular emphasis is placed on analyzing the impact of compression type and compression level on the robustness of different YOLO models. The proposed experimental framework enables a systematic comparison of performance under degraded image quality conditions and the identification of models that exhibit greater resilience to compression-induced degradation. The results of this study are expected to contribute to a better understanding of the trade-off between compression efficiency and object detection reliability, which is of particular importance in systems with limited storage and transmission resources.*

Keywords: *YOLO, object detection, image compression, DOTA, remote sensing.*

1. INTRODUCTION

Determining the exact position and class of objects in remote sensing imagery is a fundamental task in the domain of computer vision (CV). This process plays an important role in real-world applications in defense and security and can also be used for autonomous driving, environmental monitoring, updating geographic information systems (GIS), and emergency management. The application of object detection and classification algorithms to remote sensing imagery requires the use of large amounts of video content and images that need to be archived, while storage resources are limited, and transmission capacity is particularly constrained. Consequently, compression of visual signals is necessary in most cases. Lossy compression can provide several-fold savings in storage space, but its consequence is degradation of the visual quality of the image, which is unacceptable in some fields (e.g., medicine) [1], [2]. On the other hand, there are fields in which a certain degree of degradation is acceptable if it saves resources without substantially affecting object detection and recognition in the image.

Deep learning methods applied to target detection and classification in remote sensing images can generally be divided into two-stage and one-stage approaches. Two-stage detectors typically provide high detection accuracy but have higher computational requirements, which may limit real-time applications. In contrast, one-stage detectors perform object localization and classification within a unified processing framework, providing a favorable trade-off between detection accuracy and speed. The YOLO (You Only Look Once) family represents one of the most widely used groups of single-stage detectors and is known for fast and efficient real-time object detection in various application domains, including remote sensing [3]–[11].

The compression techniques analyzed in this paper

Corresponding author: N. Bukvić;
E-mail: bukvicnikol@gmail.com
DOI: 10.5937/OTEH26069B

include JPEG, which is still widely used for image compression [12], [13]. However, with the development of newer techniques such as BPG (Better Portable Graphics) [14], and especially VVC (Versatile Video Coding) [15], which represents the latest generation of video compression, the question arises as to what extent these modern approaches can improve compression efficiency in specific domains such as remote sensing. New formats introduce advanced prediction, transformation, and quantization algorithms, resulting in greater compression capability but also higher computational complexity.

Algorithm performance is evaluated using standard quantitative metrics: precision, recall, mAP, and mAP_{50}–mAP_{95} for detection and classification [16], and PSNR [17] alongside bits per pixel (bpp) for compression efficiency. Section 2 describes the image dataset, compression techniques, YOLO models and evaluation metrics. Section 3 presents and discusses the experimental results, while Section 4 summarizes the main conclusions and outlines directions for future research.

2. DESCRIPTION OF THE IMAGE DATASET, COMPRESSION TECHNIQUES, DETECTION AND CLASSIFICATION ALGORITHMS AND APPLIED MEASURES

2.1. Image dataset

The DOTA v1.0 image dataset [18] was used in this paper. The dataset contains 2,806 images with resolutions ranging from 800 to 4,000 pixels, originating from Google Earth and the GF-2 and JL-1 satellites. Most images are in color (RGB), while grayscale images were not considered. The dataset is divided into training (1/2), validation (1/6), and testing (1/3) subsets. The original images were split into 1024 × 1024 pixel images, resulting in 15,749 images divided into the same three subsets. Previously trained YOLO models from [19] were used, with 7,874 images in the training subset. From the newly formed test subset containing 5,356 images, 90 images with an average bpp value of 12.88 were selected for comparison of compression methods and detection models. The DOTA v1.0 dataset contains 15 object classes, of which 14 are represented in the selected set and were considered in the analysis. Compression-induced degradation may reduce the overlap between predicted and reference bounding boxes, leading to lower detection accuracy. The 90 images were selected to represent the test set in terms of image complexity. Spatial Information (SI_{mean}) indicates the amount of detail in an image [20], while colorfulness (CF) indicates the richness of colors [21].

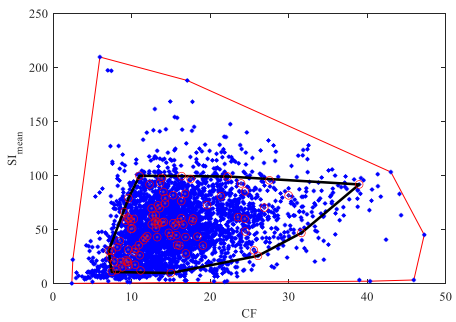


Figure 1 Relationship between colorfulness and spatial information for the DOTA v1.0 test-image subset and the selected image set for compression.

Fig. 1 shows the relationship between CF and SI_{mean} for the DOTA v1.0 test-image subset and the selected set of 90 images. In this representation, each point corresponds to one test image. Blue symbols (*) correspond to the DOTA v1.0 test-image subset (excluding grayscale images), while the quantitative indicators for the 90 selected images are shown in red. The red contour encloses all values for the original subset, whereas the black contour encloses all values for the images selected for compression. It can be observed that the black contour covers a large percentage of the area of the test-image subset, confirming that the 90 selected images are representative for further analysis of compression standards.

2.2. Compression techniques

Three compression techniques were considered in this paper: JPEG, BPG, and VVC. JPEG is a lossy image compression technique based on the discrete cosine transform [12]. Its compression strength is determined by the quality factor QF (Quality Factor), which ranges from 1 to 100, with lower QF values corresponding to lower image quality. In the experiments, QF values of 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, and 95 were used. JPEG

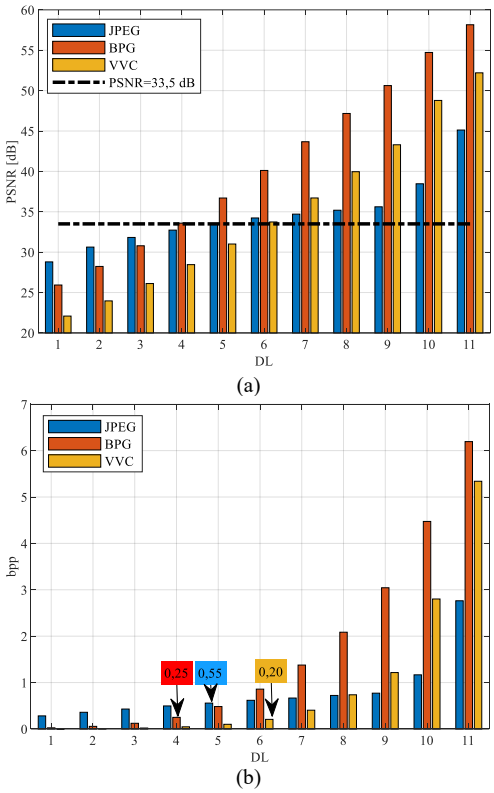


Figure 2 Objective indicators of compression techniques:
(a) PSNR and (b) bpp.

compression was performed in MATLAB using the built-in imwrite function.

Unlike JPEG, BPG is an image format based on HEVC (High Efficiency Video Coding) intra coding, developed by the JCT-VC (Joint Collaborative Team on Video Coding) [22]. The degree of compression is controlled by the parameter Q, whose values range from 0 to 51; increasing Q results in greater image degradation. The analyzed images were encoded using Q values of 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50. Compression was carried out using the BPGconv application [23] with compression level 9 (slow), chroma format 4:4:4, YCbCr color space, 8-bit depth, and the JCTVC HEVC encoder.

VVC, also known as H.266, represents the latest video-compression standard and succeeds HEVC [15]. It was primarily developed for efficient transmission of UHD video and advanced media formats such as 360° video. In VVC, compression is governed by the quantization parameter QP, which ranges from 0 to 63, with higher QP values leading to lower reconstructed-image quality. The selected QP values were 0, 10, 20, 25, 30, 35, 40, 45, 50, 55, and 60. VVC-compressed images were generated using the VVC Test Model VTM-23.11 [24] with a 4:4:4 chroma format.

2.3. Detection and classification algorithms

Seven models from the YOLO family were considered in this paper: YOLOv5, YOLOv7, YOLOv8, YOLOv9, YOLOv10, YOLOv11, and YOLOv12. YOLOv5 [5], introduced in 2020, was developed within the Ultralytics platform and is characterized by a good trade-off between detection speed and accuracy. YOLOv7 [6] introduces the trainable bag-of-freebies approach to improve performance without a significant increase in inference cost. YOLOv8 [7], introduced in 2023, uses an anchor-free detection approach and an improved architecture. YOLOv9 [8] introduces Programmable Gradient Information (PGI) and the GELAN architecture for more efficient training and preservation of important information. YOLOv10 [9] uses an end-to-end approach that enables detection without the need for the standard NMS procedure during inference. YOLOv11 [10] represents the next generation of Ultralytics models, with improved computational efficiency and feature extraction. YOLOv12 [11] integrates attention mechanisms more extensively while preserving the efficiency required for real-time detection.

Figure 3 (a) original image, (b) JPEG (QF=30), (c) BPG (Q=35) and (d) VVC (QP=35).

2.4. Applied measures

For successful comparison of different YOLO models depending on the compression techniques used, appropriate measures are required. The performance of object detection and classification algorithms was evaluated using precision (P), recall (R), mean Average Precision (mAP) calculated over IoU thresholds ranging from 0.50 to 0.95 (mAP_{50}–mAP_{95}), and mean Average Precision at an IoU threshold of 0.50 (mAP_{50}) [16]. The quality of the compressed images was evaluated using the traditional Peak Signal-to-Noise Ratio (PSNR), which is based on the mean squared error [17], while compression efficiency was expressed by the number of bits per pixel (bpp), which represents the number of bits used to represent one image pixel. It is calculated as the ratio of image size in bits to the number of pixels in the image.

3. RESULTS AND DISCUSSION

The comparison of YOLO models for object detection in compressed remote sensing images was performed in two ways. In the first approach, the objective quality measure of the compressed images, PSNR, was equalized for all three compression techniques, after which the YOLO models were compared using the previously described parameters for evaluating the performance of object detection and classification algorithms; the efficiency of the compression techniques was also presented. In the second approach, the robustness of different YOLO models to the applied compression techniques was analyzed by monitoring the mAP and mAP_{50} parameters.

3.1. Analysis for equalized PSNR values

The values shown in Fig. 2 were obtained by averaging the PSNR and bpp values over all 90 images for each of the three compression techniques and eleven representative distortion levels. The results are presented as a function of the distortion level (DL), which corresponds to the previously defined QF, Q, and QP parameter values for JPEG, BPG, and VVC, respectively. Lower DL values correspond to a higher degree of image degradation, while higher DL values represent a lower level of distortion.

It can be observed that the PSNR level for the different compression techniques is approximately the same and equals 33.5 dB for the following DL values. For JPEG compression, DL=5 (QF=30); for BPG compression, DL=4 (Q=35); and for VVC compression, DL=6 (QP=35). Fig. 2(b) shows the bpp values corresponding to the PSNR values in Fig. 2(a). For the selected PSNR level of approximately 33.5 dB, the bpp values are 0.55 for JPEG, 0.25 for BPG, and 0.20 for VVC. Consequently, at a matched objective visual quality of 33.5dB, BPG and VVC reduce the bits per pixel requirements by factors of approximately 2.2 times and 2.75 times relative to JPEG, respectively.

Fig. 3 shows an original image from the newly formed dataset (image No. 64) and the lower-right regions of its compressed versions whose values were analyzed in the previous paragraph. Degradation of the source signal can be observed in the region containing soccer fields and the surrounding grass surfaces. The blocking effect is dominant in the JPEG-compressed image, Fig. 3(b), while blurring is the dominant effect in the BPG- and VVC-compressed images, Figs. 3(c) and 3(d). In addition, it can be observed that details of the soccer-field lines are lost when BPG and VVC compression are applied.

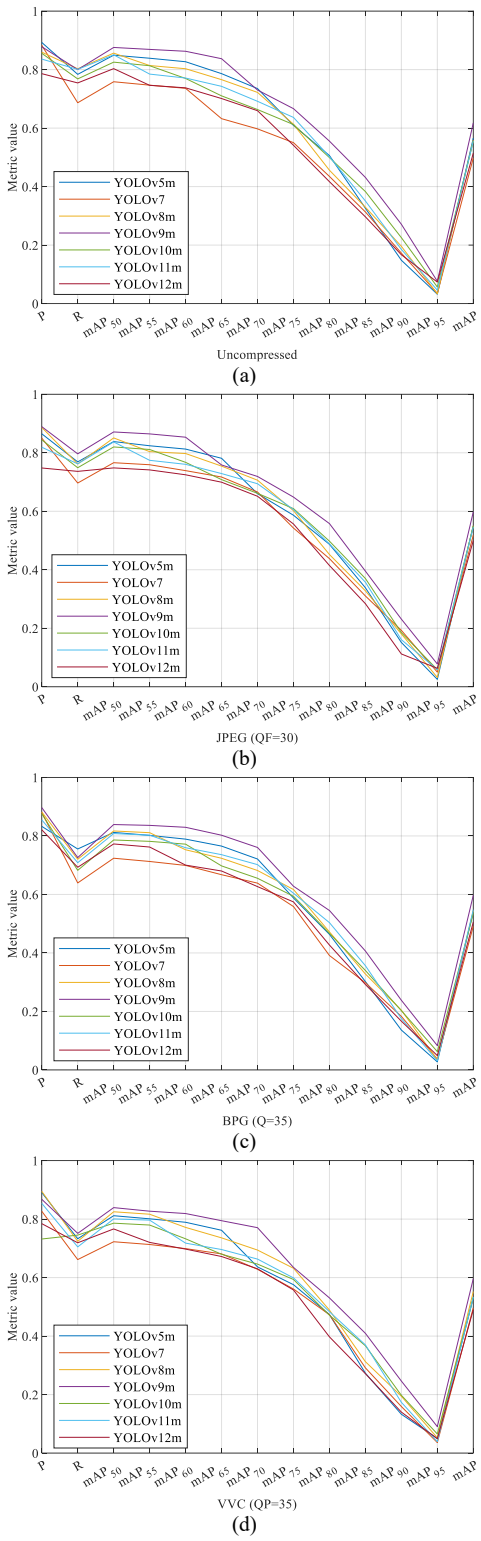
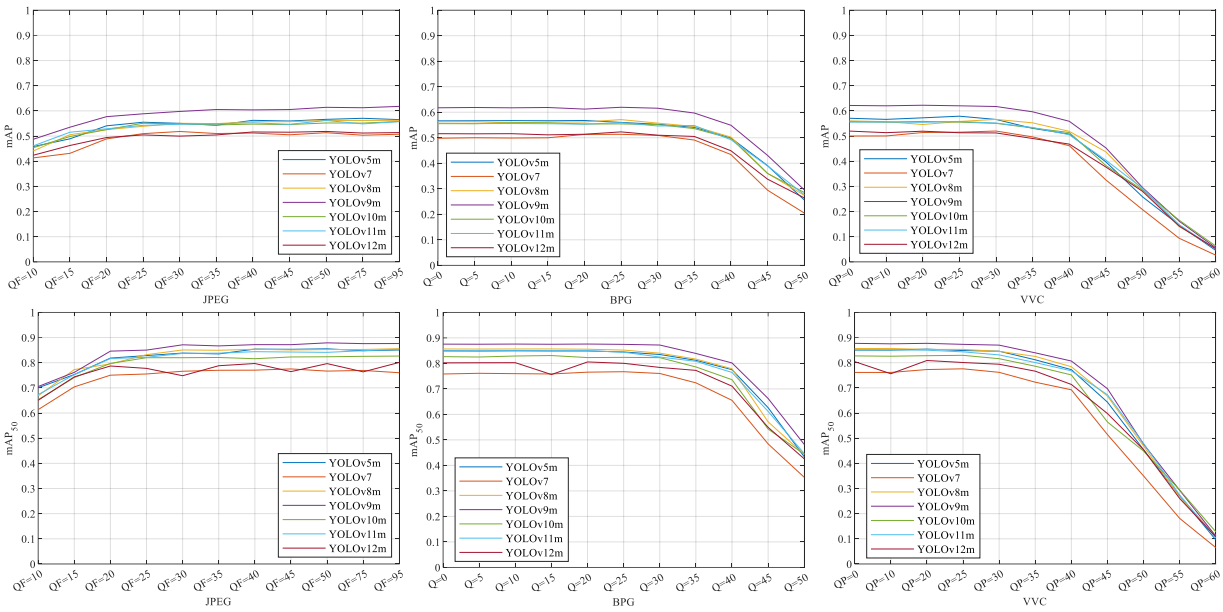


Figure 4. Characteristics of YOLO models: (a) uncompressed images, and compressed images (b) JPEG (QF=30), (c) BPG (Q=35) and (d) VVC (QP=35).



(a) (b) (c)

Figure 5 Values of the mAP and mAP_{50} parameters of YOLO models with respect to the compression techniques used: (a) JPEG, (b) BPG and (c) VVC.

Fig. 4 compares the characteristics of all seven YOLO models used on uncompressed images and on images compressed using JPEG (QF = 30), BPG (Q = 35), and VVC (QP = 35), through precision, recall, $mAP_{50\text{-}95}$ and overall mAP. For all models, mAP decreases as the IoU threshold increases, which is expected because a higher threshold requires greater precision. Fig. 4(a) shows that YOLOv9 has the highest values of the presented parameters among all applied models on uncompressed images, with an advantage of 0.01 in favor of YOLOv5 for P and mAP_{70}. The mean Average Precision (mAP) is 62%, and the mean Average Precision at an overlap threshold of 0.5 (mAP_{50}) is 88%. Fig. 4(b) shows that YOLOv9 maintains top-tier detection performance across all models on images compressed using JPEG with QF = 30, with an advantage of 0.01 in favor of YOLOv5 for mAP_{65}. The mAP is 60%, and mAP_{50} is 87%. Fig. 4(c) shows that YOLOv9 consistently dominates among all models using BPG with Q = 35, with an advantage of 0.03 in favor of YOLOv5 for R. The mAP is 60%, and mAP_{50} is 84%. Fig. 4(d) also shows that YOLOv9 achieves the highest overall accuracy on images compressed using VVC with QP = 35, with an advantage of 0.02 in favor of YOLOv5 and YOLOv8 for P. The mAP is 60%, and mAP_{50} is 84%.

3.2. Robustness analysis of YOLO models

Fig. 5 shows the mAP and mAP_{50} values of the analyzed YOLO models for JPEG, BPG, and VVC compression. At the lowest compression levels (JPEG QF = 95, BPG Q = 0, and VVC QP = 0), the obtained mAP and mAP_{50} values are approximately equal to those for uncompressed images. Among the analyzed models, YOLOv9 achieves the highest baseline values, with mAP = 0.62 and mAP_{50} = 0.88.

Fig. 5(a) shows the dependence of mAP and mAP_{50} for the YOLO models on the compression coefficient for the JPEG technique. For all YOLO models, the mAP values remain relatively stable as the compression level increases down to QF = 20, with only small fluctuations, and the reduction for individual models does not exceed 0.04. With a further decrease in QF, a more pronounced decrease in mAP is observed. Likewise, for JPEG compression, the mAP_{50} values remain relatively stable down to QF = 20 for all applied YOLO models except YOLOv12. The reduction does not exceed 0.03 for most models, while for YOLOv8 it reaches 0.06. The mAP_{50} curve for YOLOv12 exhibits somewhat different behavior, with more pronounced fluctuations at QF values of 75, 45, and 30.

Fig. 5(b) shows the dependence of mAP and mAP_{50} for the YOLO models on the compression coefficient for the BPG technique. Across all YOLO architectures, mAP metrics remain consistently stable up to quantization level Q = 30, exhibiting only minor fluctuations. For Q values above 30, a more pronounced decrease in mAP is observed for all YOLO models. Relative to the uncompressed images, mAP degradation is negligible for YOLOv7 through YOLOv9, equals 0.01 for YOLOv10–YOLOv12, and equals 0.02 for YOLOv5. Similarly, the mAP_{50} values remain resilient up to Q = 30, with overall degradation bounded within 0.03 across most models. However, the mAP50 curve for YOLOv12 localized instability, characterized by a sharp fluctuation at Q = 15.

Fig. 5(c) shows the dependence of mAP and mAP_{50} for the YOLO models on the compression coefficient for the VVC technique. For all YOLO models, the mAP values remain relatively stable for QP values up to 30, with only minimal variance. With a further increase in QP, a more pronounced decrease in mAP is observed for all YOLO models. Compared with uncompressed images, the reduction in mAP is absent for YOLOv5 and YOLOv7–YOLOv9, while it equals 0.01 for YOLOv10–YOLOv12. Similarly, the mAP_{50} values remain relatively stable as compression increases up to QP = 30, with only small fluctuations, and the reduction for individual models does not exceed 0.02. The mAP_{50} curve for YOLOv12 exhibits

somewhat different behavior due to a more pronounced fluctuation at QP = 10. At QP values of 55 and 60, corresponding to the highest VVC compression levels, YOLOv10 achieves higher mAP and mAP_{50} values than YOLOv9, whereas YOLOv9 achieves the highest values of these parameters in all other cases.

4. CONCLUSION

This work evaluated the impact of image compression standards (JPEG, BPG, and VVC) on deep learning-based object detection in remote sensing imagery using the YOLO framework. The experimental results demonstrate that modern coding standards, particularly BPG and VVC, can significantly reduce the amount of data while causing only a minimal decrease in YOLO model performance. YOLOv9 consistently achieved the highest absolute mAP and mAP50 values in most of the analyzed cases.

Future research may focus on the analysis of a larger number of images and other datasets, as well as on determining the optimal trade-off between compression level, image quality, and detection accuracy for different object classes.

References

[1] HUSSAIN,A.J., AL-FAYADH,A., RADI,N.: *Image compression techniques: A survey in lossless and lossy algorithms*, Neurocomputing, 300 (2018) 44–69, doi: 10.1016/j.neucom.2018.02.094.

[2] AL-JARRAH,O.Y., YOO,P.D., MUHAIDAT,S., KARAGIANNIDIS,G.K., TAHA,K.: *Efficient machine learning for big data: A review*, Big Data Research, 2(3) (2015) 87–93, doi: 10.1016/j.bdr.2015.04.001.

[3] DIWAN,T., ANIRUDH,G., TEMBHURNE,J.V.: *Object detection using YOLO: challenges, architectural successors, datasets and applications*, Multimedia Tools and Applications, 82(6) (2023) 9243–9275, doi: 10.1007/s11042-022-13644-y.

[4] REDMON,J., DIVVALA,S., GIRSHICK,R., FARHADI,A.: *You Only Look Once: Unified, Real-Time Object Detection*, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[5] JOCHER,G.: *Ultralytics YOLOv5*, 2020, [Online]. Available: https://github.com/ultralytics/yolov5, doi: 10.5281/zenodo.3908559.

[6] WANG,C.-Y., BOCHKOVSKIY,A., LIAO,H.-Y.M.: *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464–7475, doi: 10.1109/CVPR52729.2023.00721.

[7] JOCHER,G., CHAURASIA,A., QIU,J.: *Ultralytics YOLOv8*, version 8.0.0, 2023, [Online]. Available: https://github.com/ultralytics/ultralytics.

[8] WANG,C.-Y., YEH,I.-H., LIAO,H.-Y.M.: *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*, 2024, [Online]. Available: https://arxiv.org/abs/2402.13616.

[9] WANG,A., CHEN,H., LIU,L., CHEN,K., LIN,Z., HAN,J., DING,G.: *YOLOv10: Real-Time End-to-End Object Detection*, 2024, [Online]. Available: https://arxiv.org/abs/2405.14458, doi: 10.48550/arXiv.2405.14458.

[10] JOCHER,G., QIU,J.: *Ultralytics YOLO11*, version 11.0.0, 2024, [Online]. Available: https://github.com/ultralytics/ultralytics.

[11] TIAN,Y., YE,Q., DOERMANN,D.: *YOLOv12: Attention-Centric Real-Time Object Detectors*, Advances in Neural Information Processing Systems, 38 (2025), doi: 10.52202/085713-2627.

[12] WALLACE,G.K.: *The JPEG still picture compression standard*, Communications of the ACM, 34(4) (1991) 30–44, doi: 10.1145/103085.103089.

[13] WANG,Y., MUKHERJEE,D.: *The discrete cosine transform and its impact on visual compression: Fifty years from its invention [Perspectives]*, IEEE Signal Processing Magazine, 40(6) (2023) 14–17, doi: 10.1109/MSP.2023.3282775.

[14] BONDŽULIĆ,B.P., STOJANOVIĆ,N.M., LUKIN,V.V., KRYVENKO,S.S.: *Visually lossless image compression: From JPEG to BPG and vice versa*, IEEE Signal Processing Letters, 32 (2025) 2942–2946, doi: 10.1109/LSP.2025.3587601.

[15] INTERNATIONAL TELECOMMUNICATION UNION: *Recommendation ITU-T H.266 (V3) (09/2023): Versatile video coding*, International Telecommunication Union, Geneva, Switzerland, 2023.

[16] PADILLA,R., NETTO,S.L., DA SILVA,E.A.B.: *A Survey on Performance Metrics for Object-Detection Algorithms*, 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.

[17] WANG,Z., BOVIK,A.C.: *Mean squared error: Love it or leave it? A new look at signal fidelity measures*, IEEE Signal Processing Magazine, 26(1) (2009) 98–117, doi: 10.1109/MSP.2008.930649.

[18] DING,J., XUE,N., XIA,G.S., BAI,X., YANG,W., YANG,M.Y., BELONGIE,S., LUO,J., DATCU,M., PELILLO,M., ZHANG,L.: *Object detection in aerial images: A large-scale benchmark and challenges*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11) (2022) 7778–7796, doi: 10.1109/TPAMI.2021.3117983.

[19] ADLI,T., BUJAKOVIĆ,D., BONDŽULIĆ,B., LAIDOUNI,M.Z., ANDRIĆ,M.: *Robustness of YOLO models for object detection in remote sensing images*, Journal of Electrical Engineering, 76(5) (2025) 429–442, doi: 10.2478/jee-2025-0045.

[20] YU,H., WINKLER,S.: *Image complexity and spatial information*, Proceedings of the IEEE 5th

International Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Austria, 2013, pp. 12–17, doi: 10.1109/QoMEX.2013.6603194.

[21] HASLER,D., SUESTRUNK,S.E.: *Measuring colourfulness in natural images*, Proceedings of Human Vision and Electronic Imaging VIII, SPIE, 5007 (2003) 87–95, doi: 10.1117/12.477378.

[22] BELLARD,F.: *BPG image format*, [Online]. Available: https://bellard.org/bpg/, accessed: Apr. 01, 2026.

[23] *BPGconv 2.5*, [Online]. Available: https://bpgconv.software.informer.com/, accessed: Apr. 01, 2026.

[24] *VTM Reference Software for VVC*, [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM, accessed: Apr. 01, 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

STRUCTURAL INTEGRITY AND FEM-BASED OPTIMIZATION OF A MISSILE FIN SERVO MECHANISM

NEBOJŠA JOVIČIĆ
Military Technical Institute, Belgrade, nesapz@yahoo.com

MARKO BOJANIĆ
Military Technical Institute, Belgrade, marko.bojanic@06.vti

ALEKSANDAR STEFANOVIĆ
Military Technical Institute, Belgrade, stefanovicva@outlook.com

MARIJANA STOJANOVIĆ
Military Technical Institute, Belgrade, marijana552.17@gmail.com

Abstract: *This study extends previous research on DC motor-driven servo mechanisms for missile fin control by incorporating finite element method (FEM) analysis of the moving components. The maximum aerodynamic moment acting on the missile fin was adopted as the primary operational load for the structural assessment. A detailed three-dimensional model of the servo mechanism was developed, and a static structural analysis was performed to identify critical sections and evaluate the mechanical integrity of key components. Stress, strain, and deformation distributions were quantified under operational loading conditions, revealing potential failure zones and the most critical regions of the mechanism. The highest stress concentrations were identified at the actuator arm and transmission linkage, while maximum deformation occurred in the output lever assembly. The obtained FEM results provide a basis for future design modifications and structural optimization aimed at improving reliability and stiffness while maintaining functional performance.*

Keywords: *FEM analysis, servo mechanism, brushed DC motor, missile fin control surfaces.*

1. INTRODUCTION

Missile control systems require compact, lightweight, and highly reliable actuation mechanisms capable of generating sufficient torque for accurate positioning of aerodynamic control surfaces under demanding flight conditions. During high-speed manoeuvres, aerodynamic hinge moments acting on missile fins produce significant mechanical loads that are transmitted through the actuator mechanism. Consequently, structural stiffness, strength, and reliability of the mechanical transmission system are essential requirements for ensuring accurate fin positioning and safe operation throughout the missile flight envelope.

Electromechanical actuators (EMAs) are increasingly employed in modern aerospace applications owing to their compact architecture, high efficiency, reduced maintenance requirements, and improved reliability compared with conventional hydraulic actuation systems [1]. Recent research has focused on the design and development of aerospace electromechanical actuators through improved numerical modelling, virtual verification techniques, and optimization strategies aimed at enhancing actuator performance and supporting engineering design [2–5]. In particular, recent studies have investigated numerical modelling approaches for aerospace electromechanical actuators [2], virtual ground testing methodologies for actuator performance evaluation [3], modelling and identification of electromechanical actuators for rocket applications [4], and optimization of missile actuation systems [5]. Furthermore, fatigue assessment is becoming increasingly important during the verification and certification of aerospace electromechanical actuators [6].

Despite these developments, comparatively less attention has been devoted to the structural assessment of the mechanical transmission system responsible for transmitting the aerodynamic loads from the control surface to the actuator. In particular, finite element investigations focusing on stress distribution, deformation, and structural safety of missile fin transmission mechanisms subjected to maximum aerodynamic hinge moments remain relatively limited in the available literature. Structural verification at the design stage is therefore essential for identifying critical load-bearing components, evaluating the available safety margin, and providing a basis for subsequent structural optimization.

The present study builds upon the authors' previous research on the design and control of DC motor-driven electromechanical fin actuators [7] and on the mechanical design of a servo mechanism for controlling missile fins in the pitch and yaw planes [8]. While the earlier studies

Corresponding author: N. Jovičić;
E-mail: nesapz@yahoo.com
DOI: 10.5937/OTEH26070J

addressed actuator control, kinematic design, and functional characteristics of the actuation system, the present paper focuses on evaluating the structural integrity of the mechanical transmission system using the finite element method (FEM). A three-dimensional finite element model was developed to evaluate stress distribution, deformation, and the most highly loaded structural components under the maximum aerodynamic hinge moment acting on the missile fin. The obtained results provide engineering guidance for future structural optimization while preserving the required strength, stiffness, and positioning accuracy.

2. SERVO MECHANISM DESCRIPTION

The servo mechanism analyzed in this study was developed for the actuation of missile control fins in the pitch and yaw planes. Its primary function is to convert the rotational motion of a brushed DC motor into the angular displacement of the control surfaces while providing sufficient torque and positioning accuracy under demanding aerodynamic loading conditions. The complete design procedure and kinematic analysis of the mechanism were presented in the authors' previous work [7, 8].

The actuation system consists of a brushed DC motor, a reduction gearbox, an actuator linkage, transmission shafts, crank elements, and output levers connected to the missile fins. Motion is transmitted from the motor through the mechanical transmission system to the output shafts, producing synchronized rotation of the control fins. The mechanism is designed to provide high structural stiffness while maintaining compact dimensions compatible with the available installation space inside the missile body.

Since the actuator is intended for operation during high-speed flight, the mechanical transmission is subjected to considerable aerodynamic hinge moments generated on the missile fins. These loads are transferred through the actuator linkage and crank mechanism, producing complex stress states in several structural components. Consequently, the mechanical integrity of the transmission system represents one of the principal design requirements.

For the purpose of structural analysis, a simplified three-dimensional CAD model was created by eliminating geometrical features that have negligible influence on the global stress distribution, such as small fillets, chamfers, and fastening details. This simplification significantly reduced computational cost while preserving the structural behaviour of the mechanism under operational loading. The simplified geometry used for the finite element analysis is shown in Figure 1.

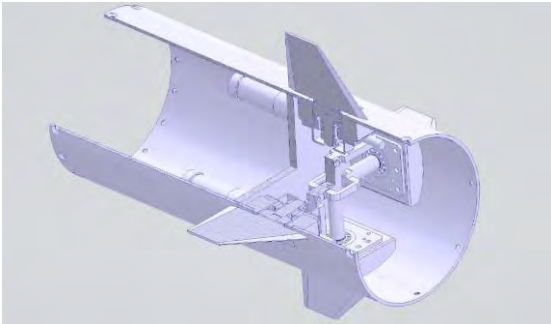


Figure 1. Cross-sectional view of the mechanical transmission system

The maximum aerodynamic hinge moment acting on the missile fin, equal to 15 Nm, was adopted as the governing design load for the structural assessment. This loading condition corresponds to the most demanding operational case considered during the design process and provides the basis for evaluating the strength, stiffness, and structural safety of the servo mechanism.

3. FINITE ELEMENT MODEL

To evaluate the structural integrity of the missile fin servo mechanism under operational loading conditions, a finite element model (FEM) was developed based on a simplified three-dimensional CAD assembly of the mechanical transmission system. The original geometry was simplified by removing small geometric features, such as fillets, chamfers, and fastening details, which have a negligible influence on the global structural response, while considerably increasing computational effort. The resulting model retained all load-carrying components essential for accurately representing the stress distribution within the transmission system.

The mechanical transmission system of the servo mechanism consists of transmission shafts, crank elements, actuator bushings, and output levers responsible for transferring torque from the brushed DC motor to the missile control fins. All structural components considered in the analysis were assumed to be manufactured from high-strength alloy steel Č.4732, characterized by a tensile strength of 1380 MPa, a yield strength of 1240 MPa, and a Young's modulus of elasticity of 210 GPa. These material properties were adopted throughout the numerical analysis.

The finite element analysis was performed using the MSC Nastran solver. The numerical model consisted of 508,663 eight-node three-dimensional hexahedral (Hex8) solid elements and 603,201 nodes. The adopted mesh density provided an appropriate representation of the structural response of the mechanical transmission system while maintaining acceptable computational efficiency. The interfaces between adjacent structural components were defined as bonded contacts, assuming perfect mechanical connection and no relative displacement between mating surfaces. The generated finite element mesh is presented in Figure 2.

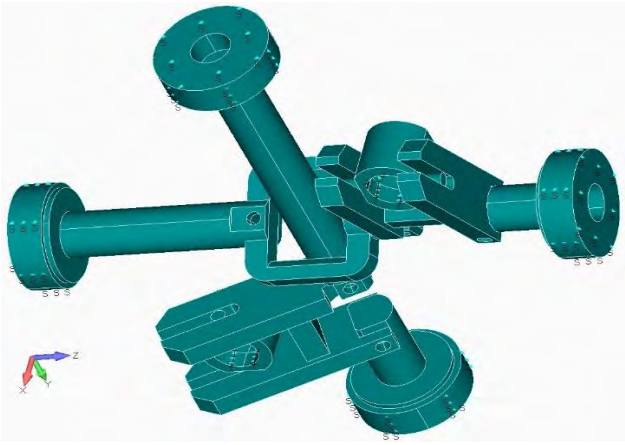


Figure 2. Finite element model of the mechanical transmission system

The applied boundary conditions were selected to represent the actual operating configuration of the mechanical transmission system. The transmission shafts were supported by bearings allowing free rotation while preventing axial displacement, whereas the actuator bushing was constrained in the operating position corresponding to a fin deflection angle of 25°.

The external loading was represented by the maximum aerodynamic hinge moment acting on the missile fin. A design torque of 15 Nm was applied to the fin shaft, representing the most severe loading condition considered during the design process. Under these assumptions, a linear static finite element analysis was performed to determine the equivalent (von Mises) stress distribution, total deformation, and critical load-bearing regions of the mechanical transmission system.

4. RESULTS AND DISCUSSION

The structural performance of the mechanical transmission system was evaluated by means of a linear static finite element analysis under the maximum operational load corresponding to an aerodynamic hinge moment of 15 Nm. The obtained stress and deformation fields provide valuable insight into the load transfer mechanism and the structural behaviour of the servo assembly.

The distribution of the equivalent (von Mises) stress for the complete mechanical transmission system is presented in Figure 3. The analysis indicates that the highest stress concentration occurs in the crank element of the transmission system, where the maximum equivalent stress reaches 604 MPa. This location corresponds to the region where the transmitted torque changes direction, resulting in combined bending and torsional loading.

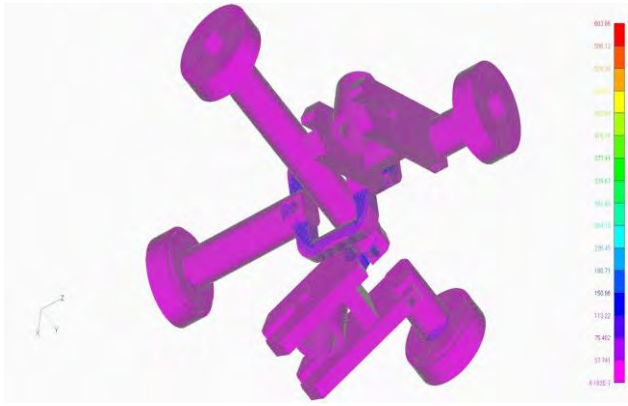


Figure 3. Equivalent (von Mises) stress distribution in the mechanical transmission system (σmax = 604 MPa)

Considering the yield strength of the selected structural material (1240 MPa), the calculated Factor of Safety (FoS) is approximately 2.05, confirming that the mechanical transmission system satisfies the structural strength requirements under the maximum design load. The obtained safety margin also indicates the existence of structural reserve, which may be utilized in future design optimization aimed at reducing the overall mass of the assembly while preserving the required mechanical performance.

A more detailed stress analysis of individual components is presented in Figures 4–6. The transmission shafts experience a maximum equivalent stress of 335 MPa, which is significantly below the allowable stress for the selected material, indicating a relatively low structural utilization. In contrast, the crank element represents the most highly loaded component, reaching the previously mentioned maximum stress of 604 MPa. The actuator bushings are subjected to substantially lower stresses, with a maximum equivalent stress of only 90 MPa, confirming that these components possess a considerable structural safety margin.

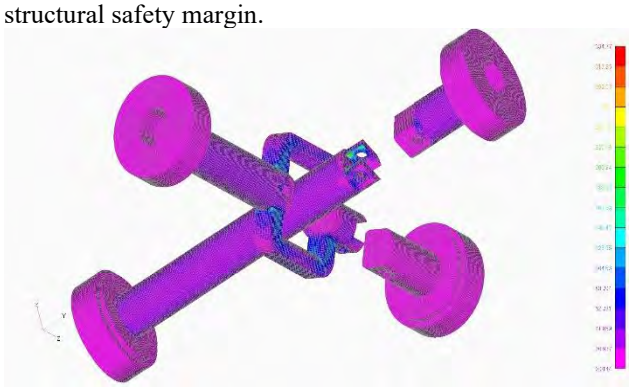


Figure 4. Equivalent (von Mises) stress distribution in the transmission shafts (σmax = 335 MPa)

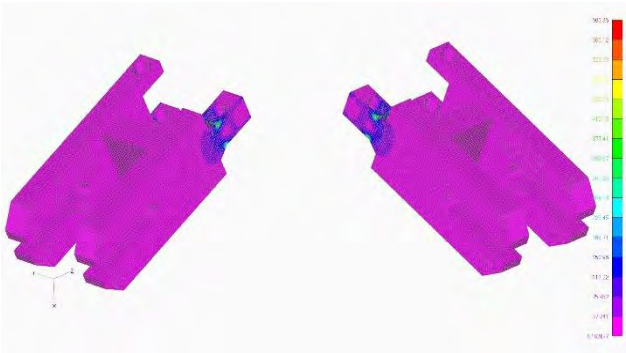


Figure 5. Equivalent (von Mises) stress distribution in the crank element (σmax = 604 MPa)

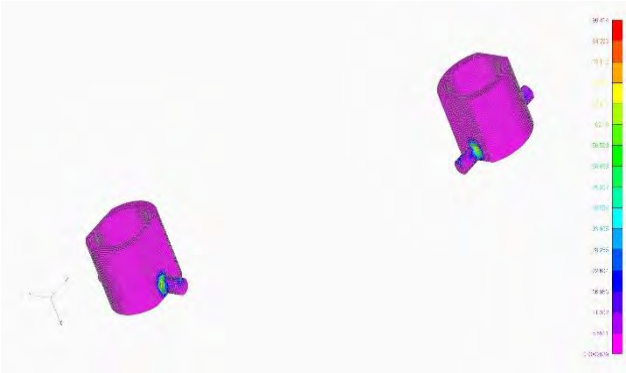


Figure 6. Equivalent (von Mises) stress distribution in the actuator bushings (σmax = 90 MPa)

The total deformation of the mechanical transmission system is shown in Figure 7. The maximum displacement amounts to 0.16 mm and occurs at the outer edge of the missile fin shaft. Such a small displacement demonstrates the high stiffness of the transmission system and ensures accurate transmission of the actuator motion to the missile control surfaces. Consequently, the positioning accuracy of the control fins is expected to be limited primarily by the actuator performance rather than by structural flexibility of the transmission mechanism.

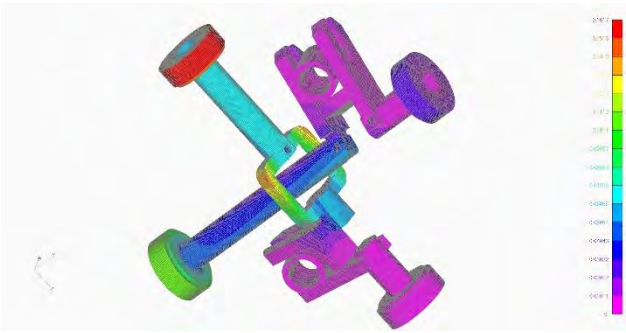


Figure 7. Total displacement of the mechanical transmission system (δmax = 0.16 mm)

Overall, the numerical results demonstrate that the proposed mechanical transmission system provides adequate structural strength and stiffness for the prescribed operating conditions. The stress distribution reveals no regions approaching the material yield limit, while the low deformation confirms the suitability of the selected structural configuration for high-precision missile fin actuation. Furthermore, the calculated safety factor suggests that the current design could be further optimized by reducing material in low-stress regions, thereby decreasing the overall mass without compromising structural integrity.

5. CONCLUSIONS

This paper presented a finite element structural assessment of the mechanical transmission system of a missile fin servo mechanism designed for pitch and yaw control. The analysis was performed under the maximum aerodynamic hinge moment acting on the missile fin, representing the most demanding operational loading condition.

The numerical results demonstrated that the proposed design satisfies the prescribed structural requirements. The maximum equivalent stress of 604 MPa was observed in the crank element, while the maximum displacement of 0.16 mm occurred at the output shaft of the missile fin. Both values remain within acceptable limits for the selected material, resulting in a calculated Factor of Safety of approximately 2.05. The relatively small deformation confirms the high stiffness of the mechanical transmission system, which is essential for accurate positioning of the missile control surfaces during operation.

The stress distribution further indicates that the transmission shafts and actuator bushings operate with considerable structural reserve, whereas the crank element represents the most highly loaded component of the assembly. These findings provide valuable guidance for future design improvements and identify the components that require particular attention during subsequent optimization.

The performed finite element analysis confirms the structural integrity of the proposed servo mechanism and provides a reliable basis for further development. Future research will focus on structural optimization aimed at reducing the overall mass of the mechanical transmission system while maintaining the required strength, stiffness, and positioning accuracy. Additional investigations will include fatigue assessment and experimental validation of the numerical model under representative operating conditions.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] ANNAZ, F. Y., KALUARACHCHI, M. M.: Progress in Redundant Electromechanical Actuators for Aerospace Applications. Aerospace, 10(9), (2023) 787.

[2] ARESCO, R.: *Advanced Numerical Models for Electromechanical Actuators*. PhD Dissertation, Politecnico di Torino, 2023

[3] YANG, S., LI, C., GUO, N., XU, C., & HU, C.: *Development of a Virtual Ground Testing Model for Aerospace Electromechanical Actuator Design Performance Evaluation*. In *International Conference on Applied Nonlinear Dynamics, Vibration and Control*, 322-336, 2024.

[4] KIERSKI, J., & GRACZYK, H.: *Modeling and Identification of Electromechanical Actuators for the ILR-33 AMBER Rocket*. *Pomiary Automatyka Robotyka*, 28(3), (2024) 53–60.

[5] TIDBALL, M., BEBLO, R., CRUZ-GONZALEZ, T., & FRANK, G.: *Optimization of an Actuation System for Articulated Missiles*. *AIAA Scitech 2022 Forum*, 2022.

[6] KIM, Y. C., KIM, D. H., KIM, S. W.: *Evaluation of the Multiaxial Fatigue Life of Electro-Mechanical Actuator for Aircraft Blade Pitch Control Based on Certification Standards*. Aerospace, 11(1) (2024).

[7] ADAMOVIĆ, P., PETRONIJEVIĆ, Z., JOVIČIĆ, N., STEFANOVIĆ, A., PAVIĆ, M.: *Lead Compensator Design for DC Motor Driven Electromechanical Fin Actuator*. Scientific Technical Review, *72*(2), (2022) 44-49.

[8] JOVIČIĆ, N., STEFANOVIĆ, A., STOJANOVIĆ, M., PAVIĆ, M.: *Design of a servo mechanism for controlling missile fins in pitch and yaw planes*. In 11th International Scientific Conference on Defensive Technologies-OTEH 2024 (No. Integrated sensor systems and robotic systems-ISSRS, (2024) pp. 289-292).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INVERSE THERMOVISION ANALYSIS OF HIGH-ENERGY PHENOMENA

DRAGAN M. KNEŽEVIĆ
Military Technical Institute, Belgrade, dragankn@gmail.com

MILAN POPOVIĆ
GS, Belgrade, popovicmilan184@gmail.com

KATARINA MIŠKOVIĆ
Military Technical Institute, Belgrade, katarinasavic@gmail.com

MILICA MARJANOVIĆ
Military Technical Institute, Belgrade, milicagoran@gmail.com

JELENA PETKOVIĆ-CVETKOVIĆ
Military Technical Institute, Belgrade, jpetkovic82@gmail.com

Abstract: *Thermovision monitoring provides an effective non-contact approach for the analysis of rapid thermal processes by enabling high-speed recording of infrared radiation response and reconstruction of temperature–time evolution. This paper presents an inverse thermovision reconstruction approach for the characterization of transient high-energy explosive processes based on radiometric analysis of fireball emission and a physics-based measurement model. Experimental validation was performed using RDX-based PBX thermobaric explosive formulations, with fireball development in static conditions in open-field tests, recorded by a high-speed infrared thermovision system operating at 1466 Hz. The reconstructed temperature-time profiles were analyzed to extract characteristic parameters, including maximum temperature, time to peak temperature, cooling rate, and integral temperature index describing the cumulative thermal response. Since thermovision systems measure radiometric signals and not the direct thermodynamic temperature, the reconstruction procedure considers emissivity, spectral response, and atmospheric transmission effects. The results demonstrate that different PBX formulations exhibit distinct thermal signatures, enabling comparative characterization based on fireball evolution and cooling dynamics. The proposed methodology confirms the potential of high-speed thermovision techniques combined with inverse reconstruction methods for quantitative non-contact analysis of transient high-energy explosion phenomena.*

Keywords: *infrared thermovision; infrared radiometry; inverse temperature field reconstruction; radiometric model; high-energy explosive phenomena; fireball dynamics; PBX explosive formulations*

1. INTRODUCTION

Thermovision technique represents one of the most important non-contact methods for spatial and temporal analysis of temperature fields and rapid thermal processes. It is based on the detection of infrared radiation emitted by the observed object and the determination of temperature-related parameters using a radiometric model that incorporates radiation source characteristics, optical system properties, atmospheric transmission, and detector response [1].

Unlike contact temperature measurement methods, thermovision systems enable non-invasive monitoring of dynamic processes with high spatial and temporal resolution, which has resulted in their application in material characterization, non-destructive evaluation, energy systems, biomedical applications, and analysis of complex thermal phenomena [1]. The physical basis of thermovision measurement relies on thermal radiation, where radiation intensity and spectral distribution depend on temperature and emission characteristics of the observed surface. However, real measurement systems deviate from the ideal black-body model due to material emissivity, detector spectral response, optical transmission characteristics, and atmospheric attenuation effects [2, 3].

For rapid thermal processes, such as high-energy explosive phenomena, temperature cannot be directly measured but must be reconstructed from the recorded infrared radiation signal. This represents an inverse thermovision problem, where the temperature field is determined from an indirectly measured detector response. Reliable reconstruction requires physically based radiation transfer models, additional constraints, and appropriate regularization methods to obtain stable and physically consistent solutions [4,5].

High-speed thermovision systems are particularly relevant for the analysis of explosive processes. Unlike nuclear explosive phenomena, where energy release is associated with nuclear transformations and intense high-energy radiation, conventional and thermobaric explosives involve rapid thermochemical reactions, heating of

Corresponding author: D. Knežević;
E-mail: dragankn@gmail.com
DOI: 10.5937/OTEH26071K

detonation products, fireball formation, and subsequent energy transfer through visible and infrared radiation [6-8].

The temperature evolution of an explosive fireball depends on chemical composition, charge density, fuel-to-oxidizer ratio, explosive geometry, and experimental conditions. Measurements of high-energy HMX-based systems have demonstrated temperature levels of approximately 3600 K during early stages of detonation product development [9]. However, the measured infrared signal is not a direct temperature value but a radiometric response affected by emissivity, spectral characteristics of the measurement system, and atmospheric propagation conditions. Therefore, accurate temperature estimation requires appropriate correction procedures and inverse thermovision reconstruction methods [10].

Modern high-speed infrared thermovision systems enable monitoring of fireball evolution, including temperature rise, maximum response, and cooling phase. The resulting temperature–time profiles provide characteristic thermovision signatures that can be used for comparative analysis of different explosive formulations [6,8,11].

The aim of this paper is to develop an inverse thermovision reconstruction approach for the characterization of transient high-energy explosive processes using high-speed infrared recording. The analysis focuses on the determination of maximum reconstructed temperature, time to peak temperature, cooling rate, and integral thermal response as physically based parameters for comparative characterization of PBX explosive formulations.

2. THERMAL-RADIATIVE PHENOMENA

2.1. Energy Balance of High-Energy Processes

Transient high-energy processes are characterized by extremely rapid changes in the internal energy state of a system, during which initially deposited energy is transformed into different energy components, including thermal, mechanical, and radiative energy. In explosive processes, this transformation occurs within a very short time interval and involves heating of detonation products, shock wave formation, expansion of gaseous products, and emission of electromagnetic radiation.

The total energy balance of the system can be expressed as:

$$E_{chem}=E_{th}+E_{mech}+E_{rad}+E_{res}$$

where E_{th} represents the thermal energy accumulated in detonation products, E_{mech} represents mechanical work and shock wave energy, E_{rad} represents energy emitted in the form of electromagnetic radiation, and E_{res} represents the residual energy component not directly included in radiometric measurement.

During the initial stage of explosive development, the dominant process is the transformation of chemical energy into the internal energy of detonation products and mechanical work:

$$Echem \rightarrow Eth+Emech$$

After reaching the maximum temperature response of the system, energy transfer mechanisms toward the surrounding environment become dominant:

$$Eres=Erad+Econv+Eexp+Emix$$

where the components represent energy losses due to thermal radiation, convection, expansion of detonation products, and mixing of products with the surrounding environment.

Thermovision systems measure only the radiative component leaving the system and do not provide direct information about the total released energy of the process. The measured signal depends on temperature, emissivity, spectral characteristics of the source, optical system properties, measurement geometry, and atmospheric attenuation [1–3].

Therefore, temperature determination from the infrared signal represents an inverse problem:

$$SIR(t) \rightarrow T(t) \rightarrow Eth(t)$$

where the temperature dynamics of the system and characteristic parameters of energy evolution are reconstructed from the measured radiometric signal.

2.2. Characteristics of Explosive Phenomena

The fundamental difference between nuclear and chemical explosive processes originates from the mechanism of energy release (Table 1). In nuclear explosions, energy is released through transformations of atomic nuclei (fission or fusion), resulting in the formation of an intense radiation field within an extremely short time interval.

From the viewpoint of thermovision technique, this difference is of fundamental importance. In nuclear phenomena, infrared emission represents a later stage of system evolution, occurring after the dominant radiation shifts toward lower-energy spectral regions. In chemical explosive processes, the infrared signal more directly corresponds to the development of the fireball temperature field and the cooling dynamics of detonation products.

Table 1. Fundamental physical differences between nuclear and conventional explosive phenomena

Parameter	Nuclear explosion	Conventional/Thermobaric explosion
Energy source	Nuclear reaction (fission/fusion)	Хемијска реакција
Characteristic temperature	10^6–10^7 K in the early radiative phase	Approximately 3000−5000 K
Dominant energy transfer mechanism	Radiative transfer	Thermochemical, radiative, and mechanical processes
Early development phase	X and γ radiation	Visible and infrared emission
Dominant spectral range	X → UV → VIS → IR	VIS + IR
Shock wave development	After the initial radiative phase	Practically simultaneous with detonation
Possibility of thermovision characterization	Limited due to early radiative effects	Direct analysis of thermal response

2.3. Temperature Reconstruction

The evolution of a high-energy source can be described through several characteristic stages. During the initial phase of nuclear processes, high-energy X- and γ-radiation dominates, rendering conventional infrared systems inapplicable because of the spectral limitations of infrared detectors.

As the fireball expands, its temperature decreases and the spectral distribution of emitted radiation gradually shifts toward the visible and infrared regions. At this stage, high-speed optical and thermovision techniques can be employed to determine:

- radiation temperature, T(t);
- fireball expansion dynamics, R(t);
- radiative energy flux.

However, the reconstructed temperature corresponds to the radiation temperature rather than the direct thermodynamic temperature of the entire source volume.

The principal limitations of infrared measurement arise from:

- the limited spectral range of the detector;
- the temporal response of the measurement system;
- the variable emissivity of the source;
- atmospheric absorption and scattering;
- the inverse nature of the temperature reconstruction problem.

Consequently, thermovision systems do not measure temperature directly; instead, they record a radiometric signal that is transformed into an estimate of the temperature field through a physics-based radiative transfer model [1-5].

2.4. Explosive Processes Characterization

For conventional and thermobaric explosives, thermovision provides an effective non-contact technique for analyzing the temporal evolution of the fireball. The temperature of detonation products typically ranges from approximately 3000 to 5000 K, depending on the chemical composition of the explosive and the conditions of the explosion [10], [11].

In this study, the explosive process is characterized by the following parameters: T_{max}, t_{max}, dT/dt, E; where T_{max} is the maximum reconstructed fireball temperature, t_{max} is the time required to reach the maximum temperature response, dT/dt denotes the cooling rate, and E is the integral energy index representing a relative measure of the overall thermal response during the observation interval.

These parameters do not represent a direct measure of the chemical energy released by the explosive. Instead, they provide a physics-based framework for comparing different high-energy systems through inverse thermovision reconstruction.

3. THERMOVISION OF EXPLOSIONS

Unlike nuclear explosive phenomena, where the initial stage of energy release is dominated by intense X- and γ-radiation, conventional and thermobaric explosives release most of their energy through thermochemical reactions, heating of detonation products, fireball formation, and subsequent radiative transfer within the visible and infrared spectral regions.

Consequently, the thermovision technique provides an effective non-contact approach for analyzing the temporal evolution of the explosive fireball, estimating its temperature regime, and reconstructing the characteristic energy-related parameters of the process [10], [11].

The temperature of detonation products for most conventional high-energy explosives typically ranges from approximately 3000 to 5000 K, depending on the chemical composition of the explosive, the fuel-to-oxidizer ratio, loading density, explosive geometry, and experimental conditions [10]. These values represent the thermodynamic characteristics of the system and should not be directly identified with the temperature reconstructed from the infrared radiometric signal.

For PBX-based explosive systems, published data indicate that the maximum fireball temperature is typically around 3600 K, representing the expected physical temperature range for comparison with and validation of the experimental results presented in this study [10], [11].

As with all radiometric measurement techniques, however, the recorded infrared signal does not represent the actual source temperature. Instead, it is the result of a thermal-radiative process governed by the source temperature, emissivity, detector spectral response, and atmospheric attenuation. Therefore, the actual fireball temperature can only be estimated using an appropriate inverse thermovision reconstruction model [1], [11].

Modern high-speed infrared thermovision systems enable time-resolved monitoring of explosive processes with high temporal resolution, allowing analysis not only of the maximum temperature but also of the complete temperature evolution of the fireball.

In the present study, the explosive process is characterized by the following parameters:

- maximum reconstructed fireball temperature, T_{max},
- time to reach the maximum temperature, t_{max},
- characteristic cooling rate, dT/dt;
- integral temperature response index,

$$E_I = \int_0^{t_{obs}} T(t)\, dt$$

which represents a relative measure of the overall thermal response during the observation period.

These parameters do not constitute a direct measure of the chemical energy released by the explosive. Instead, they provide a physics-based description of the temporal evolution of the thermal process and enable quantitative comparison of different explosive formulations using inverse thermovision reconstruction.

Table 2. Capabilities of the Thermovision Technique for Different Types of High-Energy Explosives

Explosive Type	Characteristic Temperature Range	Dominant Spectral Region	Principal Reconstructed Parameters
TNT	3000–4000 K	VIS + IR	T_{max}, cooling rate, fireball radius
PBX non-thermobaric	3200–3900 K	VIS + IR	Temperature field, cooling dynamics, energy distribution
Thermobaric mixtures	3000–5000 K	VIS + IR	Fireball duration, spatial temperature distribution, energy reconstruction

For TNT and other non-thermobarics, the temperature response originates from the short-lived emission of detonation products, whereas thermobaric formulations enable analysis of rapid temperature evolution and subsequent cooling. Unlike nuclear phenomena, these explosive systems are dominated by radiation within the visible and infrared spectral regions, making the thermovision technique suitable for monitoring the maximum temperature, cooling dynamics, fireball evolution, and the relative dynamics of thermal energy release (Table 2) [6-8].

4. INVERSE TEMPERATURE MODEL

An infrared thermovision camera does not measure the temperature of an observed object directly. Instead, it records a radiometric signal that depends on the surface temperature, material emissivity, atmospheric attenuation, the spectral characteristics of the optical system, and the temporal characteristics of image acquisition. Consequently, determining the actual temperature field constitutes an inverse problem, in which the temperature distribution is reconstructed from the measured infrared signal [1, 2].

The physical basis of the reconstruction is the law of thermal radiation. For the idealized case of total radiative emission, the Stefan–Boltzmann law can be expressed as

$$M = \varepsilon\, \sigma\, T^4$$

where M is the total radiative emittance, ε is the emissivity of the source, σ is the Stefan–Boltzmann constant, and T is the absolute temperature.

In a real measurement system, the recorded signal results from the interaction of the emitted radiation with the atmosphere, the optical system, and the detector. Therefore, temperature estimation requires the application of a correction and inverse reconstruction model, which can generally be expressed as

$$T_{max} = f(S_{IR}, \varepsilon, \tau, H(\lambda), t)$$

where S_{IR} is the measured infrared signal, ε is the emissivity, τ is the atmospheric transmittance, $H(\lambda)$ denotes the spectral response of the optical and detector system, and ttt is the acquisition time relative to the initiation of the process.

The proposed methodology is based on a physics-driven model that relates the spatial and temporal distribution of deposited energy to the resulting temperature field and the recorded infrared signal:

$$Q(x,y,z,t) \rightarrow T(x,y,z,t) \rightarrow M_\lambda(x,y,t) \rightarrow Sdet(t) \rightarrow T^*(x,y,z,t)$$

where $Q(x,y,z,t)$ denotes the volumetric heat source density, $T(x,y,z,t)$ is the temperature field, $M_\lambda(x,y,t)$ is the spectral radiance, $S_{det}(t)$ is the detector signal, and $T^*(x,y,z,t)$ is the reconstructed temperature field.

Unlike direct thermometry, temperature in thermovision is not a directly measured quantity but rather a parameter obtained through inverse thermovision reconstruction. Since this inverse problem is inherently ill-posed, its solution requires the incorporation of physical constraints, model assumptions, and appropriate regularization techniques to ensure stable and physically consistent reconstruction results [2, 9].

In addition to reconstructing the temperature field, the obtained results can be used to estimate the thermal energy loading of the investigated material. For this purpose, the accumulated specific thermal energy (thermal dose) is defined as

$$D = \int_0^t \frac{Q(x,y.z.\tau)}{\rho}\, dt$$

where D is the accumulated specific thermal energy (thermal dose) [J/kg], $Q(x,y,z,t)$ is the volumetric heat source density, i.e., the rate of energy deposition within the material [W/m^3], and ρ is the material density [kg/m^3].

Unlike conventional radiation dosimetry, this quantity represents the thermal energy dose, i.e., the amount of thermal energy accumulated per unit mass of the material

during the observation interval.

However, in explosive processes the spatial distribution of the heat source, $Q(x,y,z,t)$, cannot be determined directly from infrared measurements because the recorded signal represents the outcome of a complex thermal-radiative process. Consequently, the inverse thermovision reconstruction model is primarily employed to recover the spatiotemporal temperature evolution, while the corresponding energy-related parameters are derived from characteristic quantities extracted from the reconstructed temperature response.

5. EXPERIMENTAL RESULTS

5.1. Experimental Setup and Parameters

Experimental investigations were conducted on six explosive charges, each with a mass of 1 kg, prepared in three different PBX formulations (PBX-1, PBX-2, and PBX-3). Two independent experiments were performed for each formulation to evaluate the repeatability of the reconstructed temperature response.

The samples were designated as PBXij, where the first index (i=1,2,3) denotes the formulation type, while the second index (j=1,2) identifies the repeated measurement.

Measurements were performed using a high-speed infrared thermovision camera operating at 1466 Hz, equipped with a 50-mm focal-length lens and positioned at a distance of 60 m from the explosive charge. The resulting spatial resolution was approximately 36 mm/pixel, enabling continuous monitoring of the temporal evolution of the fireball throughout the explosive event.

From the reconstructed temperature–time histories, the following characteristic parameters were extracted:
- maximum reconstructed fireball temperature, T_{max};
- time to reach the maximum temperature, t_{max};
- characteristic cooling rate, $(dT/dt)_{cool}$;
- integral temperature response index, E_I.

These parameters were used to compare different PBX formulations based on their thermal-radiative response.

5.2. Analysis of the Temperature–Time Response

For each experiment, the following characteristics were analyzed:
- the reconstructed temperature evolution, $T=T(t)$;
- the maximum reconstructed temperature, T_{max}, corresponding to the highest temperature attained by the fireball during the explosive process;
- the cooling rate, $(dT/dt)_{cool}$, evaluated after the temperature peak as the slope of the descending portion of the temperature curve.

The cooling rate is governed by several simultaneous energy dissipation mechanisms, including:
- thermal radiation;
- convective heat transfer;
- expansion of detonation products;
- mixing with the surrounding air.

To provide an additional measure of the overall thermal response, the integral temperature response index was introduced:

$$E_I = \int_0^{t_{obs}} T(t)\,dt$$

which represents a relative measure of the accumulated temperature response during the observation interval.

It should be emphasized that E_I does not represent the actual chemical energy released by the explosive. Instead, it is an integral parameter derived from the reconstructed temperature signal, enabling quantitative comparison of different explosive formulations based on their thermal response.

In practical data processing, the discrete form was employed:

$$E_I = \sum_{i=1}^{n} T_i \Delta t$$

where T_i is the reconstructed temperature in the i-th frame, nnn is the total number of analyzed frames, and

$$\Delta t = 0{,}682\ ms.$$

5.3. Statistical Analysis

Two independent measurements were performed for each PBX formulation to evaluate the repeatability and variability of the experimental response.

The following statistical parameters were determined:
- individual values of the maximum reconstructed temperature, T_{max};
- mean temperature, $\bar{T}$;
- standard deviation, σ_T ;
- indicative confidence interval.

The mean value for each formulation was calculated as

$$\bar{T}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} T_{ij}$$

where T_{ij} denotes the maximum reconstructed temperature obtained in the i-th experiment of formulation j.

The standard deviation was calculated as

$$\sigma_{T_j} = \sqrt{\frac{\sum_{i=1}^{n_j}\left(T_{ij} - \bar{T}_j\right)^2}{n_j - 1}}$$

To estimate the uncertainty of the mean value, a 95% confidence interval (used here as an indicative interval of variation) was determined using Student's t-distribution:

$$CI_{95} = \bar{T}_j \pm t_{0,975,n_j-1}\frac{\sigma_{T_j}}{\sqrt{n_j}}$$

Table 3. Statistical Parameters of the Maximum Reconstructed Temperature for the Investigated PBX Formulations

Formulat.	Sam. 1	Sam. 2	Mean (K)	SD (K)	95% CI (K)
PBX-1	2044.38	2030.87	2037.63	9.55	±85.83
PBX-2	2057.49	3090.05	2573.77	730.13	±6559.85
PBX-3	2255.55	2101.93	2178.74	108.63	±975.95

The results presented in Table 3 reveal different levels of variability among the investigated formulations. PBX-1 exhibits the highest repeatability, whereas PBX-2 shows substantially greater dispersion owing to the large difference between the two repeated measurements. PBX-3 demonstrates an intermediate level of variability.

Considering the limited number of repeated experiments (n=2), the calculated statistical parameters should be interpreted as indicators of experimental variability rather than as definitive estimates of the underlying statistical distribution.

The reconstructed temperatures fall within the physically expected range for RDX/HMX-based high-energy explosive systems, supporting the physical consistency of the applied radiometric reconstruction model [10].

The obtained results demonstrate that the thermovision technique enables effective characterization of the temporal evolution of the fireball and extraction of parameters sensitive to differences in the chemical composition of PBX formulations.

The principal limitations of the proposed methodology arise from the variable emissivity of the fireball, the spectral response of the detector, atmospheric attenuation, and the finite spatial and temporal resolutions of the measurement system. Nevertheless, the results confirm that inverse thermovision reconstruction provides a reliable and physically meaningful approach for the comparative characterization of the thermal response of different high-energy explosive systems.

6. CONCLUSONS

The capabilities and limitations of thermovision techniques for the characterization of transient high-energy explosive processes were investigated. An inverse thermovision reconstruction approach was developed using a physics-based radiometric model that relates the measured infrared signal to the reconstructed temperature–time evolution of the fireball.

Experimental analysis of PBX formulations demonstrated that thermovision techniques enable the extraction of characteristic thermal-response parameters, including the maximum reconstructed temperature, peak-time, cooling rate, and integral temperature index. The reconstructed temperatures were within the expected range for RDX/HMX-based high-energy systems, confirming the physical consistency of the proposed model and reconstruction procedure.

The analysis of nuclear explosive phenomena showed that conventional infrared systems are limited during the initial fireball phase due to dominant high-energy X-ray and ultraviolet radiation, while thermovision-based characterization becomes applicable during later evolution stages.

The obtained results confirm that thermovision techniques combined with physics-based radiative transfer models and inverse reconstruction methods provide an effective non-contact approach for the analysis and comparative characterization of conventional high-energy explosive systems.

Acknowledgements

This work was supported by the Ministry of Science, Technological development and Innovations of the Republic of Serbia, Contract No. 451-03-34/2026-03/200325, and University of Defense, Military Academy, Proj. No. VA-TT/1/22-24.

References

[1] Xavier P. V. Maldague, Infrared Thermography: Principles and Applications, book (monograph), 2nd edition, Wiley, Hoboken, USA, 2001.

[2] Robert Siegel, John R. Howell, Thermal Radiation Heat Transfer, book (monograph), CRC Press, 5th edition, 2010.

[3] Michael F. Modest, Radiative Heat Transfer, book (monograph), 3rd edition, Academic Press (Elsevier), Amsterdam / Boston, USA, 2013.

[4] Krzysztof Grysa, Inverse Heat Conduction Problems, book (monograph), Kielce University of Technology, Kielce, Poland, 2001.

[5] Alberto Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, book (monograph), SIAM, Philadelphia, USA, 2005.

[6] D. Simić, U. Andjelić, D. Knežević, K. Savić, V. Draganić, R. Sirovatka, Lj. Tomić, Thermobaric effects of cast composite explosive of different charge mass and dimensions, Central European Journal of Energetic Materials, 2016, Vol. 13, No 1, p. 161-182.

[7] U. Anđelić, D. Simić, D. Knežević, M. Dević, Visualizing the thermal effect of thermobaric explosives, 7th International Scientific Conference on Defensive Technologies OTEH 2016, Belgrade, 6-7 October 2016. Proceedings, 288-292.

[8] D. Simić, U. Andjelić, D. Knežević, K. Mišković, S. Gaćić, S. Terzić, S. Brzić, Quantification of thermal effect of a thermobaric detonation by infrared imaging technique, 18th Seminar on New Trends in Research of Energetic Materials NTREM 2015, Pardubice, Czech Republic, April 15-17, Proceedings, 829-838, 2015.

[9] J. W. Ferguson, J. C. Richley, B. D. Sutton, E. Price, T. A. Ota, The Measured Temperature and Pressure of EDC37 Detonation Products, conference paper, AIP Conference Proceedings, Vol. 1793, Paper 030013, American Institute of Physics (AIP), Melville, NY, USA, 2017. DOI: 10.1063/1.4971471.

[10] L.-Y. Wang, H.-M. Du, H. Xu, Compensation Method for Infrared Temperature Measurement of Explosive Fireball, journal paper, Thermochimica Acta, Vol. 680, Article 178342, 2019. DOI: 10.1016/j.tca.2019.178342.

[11] M. J. Hobbs, A. Barr, S. Woolford, D. Farrimond, S. D. Clarke, A. Tyas, J. R. Willmott, High-Speed Infrared Radiation Thermometer for the Investigation of Early Stage Explosive Development and Fireball Expansion, journal paper, Sensors, Vol. 22, No. 16, Article 6143, MDPI, Basel, Switzerland, 2022. DOI: 10.3390/s22166143.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

REACH AND WORKSPACE ENVELOPE ANALYSIS OF THE SIX-DOF ARTICULATED ROBOT ARMA6

JELENA VIDAKOVIĆ
Lola Institute, Belgrade, jelena.vidakovic@li.rs

ANDRIJA DEVIĆ
Lola Institute, Belgrade, andrija.devic@li.rs

GORAN VASILIĆ
Lola Institute, Belgrade, goran.vasilic@li.rs

NIKOLA ŽIVKOVIĆ
Lola Institute, Belgrade, nikola.zivkovic@li.rs

IVANA VASOVIĆ MAKSIMOVIĆ
Lola Institute, Belgrade, ivana.vasovic@li.rs

Abstract: *This paper presents the reach and position workspace envelope analysis of the six-degree-of-freedom articulated robot ARma6 based on the solution of its forward kinematic problem. The robot reach is determined by combining Monte Carlo sampling with nonlinear optimization, while the limits of Tool Centre Point (TCP) in Cartesian space are obtained analytically using trigonometric relations. The rotational symmetry of the positional workspace is exploited to simplify the three-dimensional positional workspace's envelope determination, with two approaches considered: numerical extraction of the inner and outer radial boundaries and a geometric approach based on the circular TCP path corresponding to the effective robot reach. The obtained results characterize the reach and position workspace of the ARma6 robot and provide a basis for further evaluation of its geometry during the design process.*

Keywords: *robot, kinematics, reach, workspace, forward.*

1. INTRODUCTION

Growing customer demands and global competition are amongst the key drivers of the increasing need for integrating industrial robots with Industry 4.0 solutions [1]. Flexible robotic systems can adapt to a variety of tasks and products, making them particularly advantageous for high-mix, low-volume manufacturing, which is typical for small and medium-sized enterprises (SMEs) [1-2]. However, the wider adoption of industrial robots in SMEs remains limited by factors such as investment costs and the expertise required for robot programming [1-4]. These challenges encourage the development of more accessible robotic solutions for SMEs. In this context, the ARma6 robotic system, Fig.1, is being developed as a compact six degrees of freedom (DoFs) articulated robot supported by an extended reality (XR)-based platform for robot programming and interaction [5].

Solution of the robot forward kinematics problem (FK algorithm) is used to determine the position and orientation of the robot end-effector, i.e. frame attached to its TOOL CENTRE POINT (TCP) with respect to (w.r.t.) the BASE frame, for the input joint variables [6-7]. FK serves as the basis for robot simulation, inverse kinematics (IK) and motion planning, and workspace analysis.

Robot workspace and reach are amongst the key considerations in robot design, as they provide a quantitative measure of the range of tasks that the robot can perform. Additionally, before applying the IK algorithm, it is useful to verify whether the desired TCP position lies within the robot reach, position limits, or within its workspace, allowing the user to distinguish potential workspace-related errors from other IK failures. The workspace is often broken down into a reachable (position) workspace - the entire set of points reachable by the manipulator, and a dexterous workspace - points that the manipulator can reach with an arbitrary orientation of the end-effector [8-9].

In this paper, the FK algorithm, robot reach calculation, and position workspace envelope determination of the ARma6 articulated robot are presented. The robot reach is determined numerically using nonlinear optimization. Additionally, the limits of the TCP position in Cartesian space are obtained analytically using trigonometric identities. The methodology for the positional workspace envelope determination is presented.

This paper is divided as follows. In Section 2, the FK algorithm for the ARma6 robot is presented. The methodology for determining a robot reach and position

Corresponding author: J. Vidakovic;
E-mail: jelena.vidakovic@li.rs
DOI: 10.5937/OTEH26072V

workspace envelope is presented in Section 3, and numerical simulation results are also presented herein. Discussion is given in Section 4, while concluding remarks are given in Conclusion.

Figure 1. 3D model of the ARma6 robot

2. FORWARD KINEMATICS ALGORITHM

One of the most used methods for kinematic modeling in robotics, the Denavit-Hartenberg (DH) convention [10], is adopted in this study. Robot DH parameters, Table 1, are defined based on frames attached to robot joints, Fig. 2 [5].

Table 1. Denavit-Hartenberg parameters for ARma6

Joint i	a_i	α_i	d_i	a_i
1	0	90°	d_1	θ_1
2	a_2	0	0	$\theta_2 - 90$
3	a_2	0	0	θ_3
4	0	90°	d_4	$\theta_4 - 90$
5	0	-90°	d_5	θ_5
6	0	0	d_6	θ_6

Joint-to-joint homogeneous transformation matrix (HTM) describing coordinate frame i relative to frame i-1 using DH parameters is given as follows [11]:

$$\mathbf{A}_i = \mathrm{Rot}(z_{i-1},\theta_i)\mathrm{Trans}(z_{i-1},d_i)\mathrm{Trans}(x'_{i-1},a_i)\mathrm{Rot}(x'_{i-1},\alpha_i) \quad (1)$$

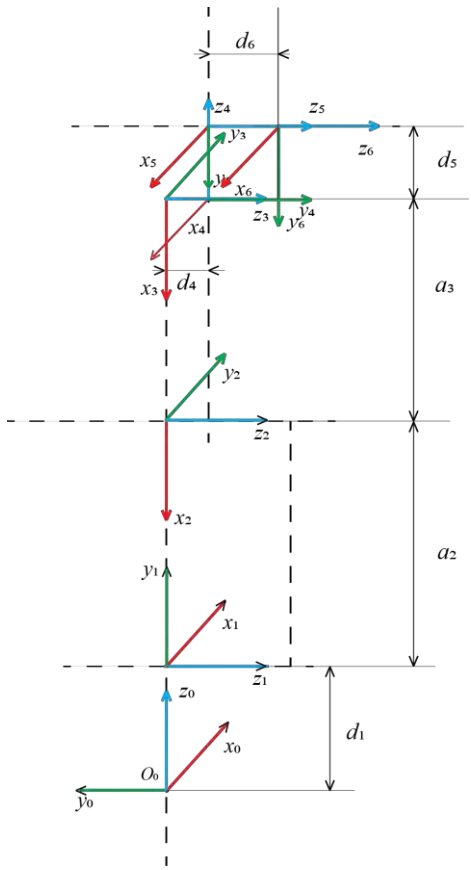


Figure 2. Frames assigned to ARma6 joints [5]

Based on joint-to-joint HTMs, the HTM of the frame attached to the robot's TCP w.r.t. the BASE frame is obtained as:

$$\mathbf{T}_{06}(\mathbf{q}) = \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\mathbf{A}_4\mathbf{A}_5\mathbf{A}_6 \quad (2)$$

In Eq. (2), $\mathbf{q} = [q_1 \ \ldots \ q_n]$ is a vector defining the robot configuration (joint positions q_i for i-1..n, where n is the number of DoFs, i.e. 6 for ARma6).

Hereinafter, the following notation is adopted: $c_i = \cos(q_i)$, $s_i = \sin(q_i)$, $c_{ij} = \cos(q_i + q_j)$, $s_{ij} = \sin(q_i + q_j)$, $c_{ijk} = \cos(q_i + q_j + q_k)$, $s_{ijk} = \sin(q_i + q_j + q_k)$

When the HMT $\mathbf{T}_{06}(\mathbf{q})$ is defined, the pose of the TCP w.r.t. the BASE frame can be calculated for every time instance t, i.e. the robot configuration $\mathbf{q}(t)$ given that $\mathbf{T}_{06}$ contains the rotation matrix and the position vector of the TCP frame and its origin w.r.t. the BASE frame [6]. X, Y, Z coordinates are obtained from the fourth column of the matrix $\mathbf{T}_{06}$ given in Eq. (2) as follows:

$$X(\mathbf{q}) = d_6(c_5 s_1 + c_{234} c_1 s_5) + d_4 s_1 - a_2 c_1 s_2 - d_5 s_{234} c_1 - a_3 c_1 c_2 s_3 - a_3 c_1 s_2 c_3 \quad (3)$$

$$Y(\mathbf{q}) = d_6(c_{234} s_1 s_5 - c_1 c_5) - a_2 s_1 s_2 - d_5 s_{234} s_1 - a_3 s_1 s_{23} - d_4 c_1 \quad (4)$$

$$Z(\mathbf{q}) = d_1 + a_3 c_{23} + a_2 c_2 + d_5 c_{234} + d_6 s_5 s_{234} \quad (5)$$

TCP frame orientation is obtained using the RPY angles convention [12] as given in [5].

3. DETERMINATION OF ROBOT REACH AND POSITION WORKSPACE ENVELOPE

Both the robot reach and the size of the workspace envelope are fundamental geometric characteristics that strongly influence the mechanical design and application of robotic manipulators. Increasing the workspace and reach typically requires longer links, which generally reduce manipulator stiffness and positioning accuracy. For the ARma6 robot and its adopted payload, a target reach of approximately 1 m was selected during the initial design stage, resulting in a compact robot configuration with a workspace suitable for the intended development and experimental activities.

3.1. Robot reach determination

In industrial robot specifications, the reach is commonly reported as the maximum distance from the robot BASE FRAME origin to the TCP, here denoted as $R_{\max}$, or as the maximum horizontal radial distance of the TCP to the BASE FRAME vertical axis, here denoted as $\rho_{\max}$.

$$R_{\max} = \max\sqrt{X^2 + Y^2 + Z^2} \quad (6)$$

$$\rho_{\max} = \max\sqrt{X^2 + Y^2} \quad (7)$$

Taking into account Eqs. (3) and (4), and by using simple trigonometric identities, the distance $\rho(\mathbf{q})$ from Eq. (7), i.e. the horizontal radial reach, can be calculated as

$$\rho(\mathbf{q}) = \sqrt{A^2 + B^2} \quad (8)$$

where

$$A = A(\mathbf{q}) = d_6 c_{234} s_5 - a_2 s_2 - d_5 s_{234} - a_3 s_{23} \quad (9)$$

$$B = B(\mathbf{q}) = d_4 + d_5 c_5 \quad (10)$$

The robot reach $R_{\max}$ can be calculated using several different approaches. Although the reach may be estimated by evaluating the FK over a discretized joint space, such an approach does not guarantee that the global maximum is found. Calculating reach for selected DH parameters can be regarded as a nonlinear optimization problem in which the joint variables $q_1,.., q_6$ are treated as optimization variables and the Euclidean distance from the robot BASE FRAME origin to the TCP is maximized. Given that the reach function is a highly nonlinear trigonometric function of multiple joint variables, obtaining a closed-form solution is generally challenging. In this study, numerical approaches are utilized.

It can be seen from Eq. (5-6) and Eqs. (8-10) that $Z(\mathbf{q})$ and $\rho(\mathbf{q})$ and consequently $R(\mathbf{q})$ do not depend on q_1 and q_6. Consequently, herein the maximum reach was determined numerically for the selected DH parameters by „optimizing“ the joint variables q_2, to q_5. The following two-stage procedure was applied for $R_{\max}(\mathbf{q})$ calculation:

1. Generate a large number of random initial configurations within the admissible joint space using Monte Carlo [13] sampling.
2. Starting from each initial configuration, locally optimize the objective function. The largest value obtained from all optimization runs is taken as the estimated global maximum reach.

On another note, since d_1 represents a fixed offset between the BASE FRAME and the second joint, it is convenient to define the effective robot reach $R_{1\max}$ w.r.t. the frame attached to the second joint, as illustrated in Fig. 3. The effective robot reach is calculated as:

$$R_{1\max} = Z_{\max} - d_1 \quad (11)$$

For calculation of $Z_{\max}$, the approach using trigonometric identities (without optimization methods) is applied herein. From Eq. (5), it can be seen that for the maximum TCP's Z coordinate w.r.t. the BASE FRAME, the robot configuration has to be the following: $q_2 = 0,\ q_2 + q_3 = 0 \rightarrow q_3 = 0,\ q_5 = \pi/2$. For the adopted values, the expression:

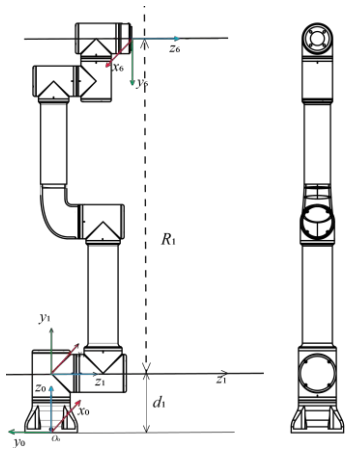


Figure 3. Robot ARma6 model selected frames.

$$d_5 c_{234} + d_6 s_{234} \quad (12)$$

has the maximum $\sqrt{d_5^2 + d_6^2}$ for the value of q_4:

$$q_4 = \alpha,\ \alpha = a\tan 2(d_6, d_5) \quad (13)$$

and the following is obtained:

$$Z_{\max} = d_1 + a_2 + a_3 + \sqrt{d_5^2 + d_6^2} \quad (14)$$

$$Z_{\min} = d_1 - a_2 - a_3 - \sqrt{d_5^2 + d_6^2} \quad (15)$$

3.2. Robot position workspace envelope determination

The workspace envelope encloses the workspace. Based on the FK algorithm, Eqs. (3-5), the position workspace can be calculated numerically. From Eq. (5) and Eqs. (8-10), it can be seen that neither the radial distance $\rho(Z)$ of the TCP from the BASE FRAME vertical axis, i.e., z_0 axis from Fig.2, nor the TCP's Z coordinate depend on the angle q_1. The joint angle q_1 changes only the azimuthal position of the TCP about the z_0 axis. From this, it can be concluded that the positional workspace of the ARma6 robot is rotationally symmetric around the BASE FRAME vertical axis.

In this study, the positional workspace envelope was determined using two approaches. In both approaches, the analysis was performed for an adopted constant value of q_1, and the obtained two-dimensional geometric representation was subsequently revolved about the z_0 axis, Fig. 2, to generate the three-dimensional workspace envelope. Approach 1 is based on the numerical determination of the inner and outer workspace boundaries in the ρ-Z plane. Approach 2 is based on the determination of the complete circular TCP path obtained by the rotation of the robot arm corresponding to the effective robot reach $R_{1\max}$, Eq. (11), Fig. 3, about the z_1 axis, $q_2 \in [0,\ 2\pi]$.

Approach 1

In Approach 1, the three-dimensional positional workspace is determined by one meridional cross-section (e.g. $q_1 = 0$). The Cartesian coordinates are transformed into cylindrical coordinates (ρ, Z). The inner $\rho_{\min}(Z)$ and outer $\rho_{\max}(Z)$ boundaries of the meridional section are determined using the following numerical procedure:

1. Perform a grid search [14] over the selected joint variables. Compute the TCP coordinates (X, Y, Z) using the FK for all sampled joint configurations and calculate the corresponding radial distance ρ, Eqs. (8-10).
2. Divide the range of Z-coordinates into a predefined number of horizontal layers of equal thickness. For each layer, identify all sampled points whose Z-coordinates lie within the corresponding interval.
3. Determine the inner and outer boundaries of all layers as the minimum and maximum values of $\rho(Z)$ among the selected points.
4. After obtaining the inner $\rho_{\min}(Z)$ and outer $\rho_{\max}(Z)$ boundaries for each layer in $Z \in [Z_{\min},\ Z_{\max}]$ as an array of discrete values, connect them by linear interpolation.
5. Revolve interpolated $\rho_{\min}(Z)$ and $\rho_{\max}(Z)$ curves about the the base-frame vertical z_0 axis.

Approach 2

The second approach in generating the outer workspace envelope is based on the complete rotation of the robot arm about the z_1 axis, $q_2 \in [0,\ 2\pi]$ in the configuration corresponding to the effective robot reach $R_1 = R_{1\max}$, for for an arbitrary adopted value of q_1 angle (herein it is adopted that $q_1 = 0$). The circular TCP path can be obtained either analytically, using the parametric equation of a circle with the previously determined center and radius $r = R_{1\max}$, or numerically, using the FK for a complete variation of the corresponding joint angle $q_2 \in [0,\ 2\pi]$. In both cases, the obtained TCP coordinates are subsequently transformed by the rotation matrix about the BASE FRAME vertical axis. The procedure consists of the following steps:

1. Determine the effective robot reach $R_{1\max}$, the corresponding robot configuration $q_{2\ldots}q_5$, and the center C_0 of the resulting TCP circle.
2. Generate the circular TCP path either from its parametric equation or directly from the FK by varying the corresponding joint angle through a complete revolution.
3. Revolve the obtained circular TCP path about the BASE FRAME vertical axis by applying the corresponding rotation matrix to the TCP position vectors $[\mathbf{X\ Y\ Z}]^T$.

3.3. Numerical results

For numerical calculation, the following DH parameters have been adopted: d_1=158.09mm, d_4=101mm, d_5=86mm, d_6=90.29mm, a_2=158.09mm, a_3=158.09mm.

For the selected DH parameters, 50000 random initial configurations of the joint variables q_2, q_3, q_4, q_5 were generated using Monte Carlo sampling within the range [0, 2π]. Starting from each initial configuration, a constrained local optimization was performed in Matlab [15]. The maximum reach obtained was $R_{\max} = 110.54$cm for $q_2 \approx$ -0.03°, $q_3 \approx$ 0.02°, q_4= -46.11°, q_5=-81.83°.

From Eq. (14-15), based on adopted DH parameters, it is obtaoned that $Z_{\max}$ =110.02 cm, $Z_{\min}$ =-78.4 cm. From Eq. (13) and Eq. (11), it is obtained that α=46.39°, and $R_{1\max} = 94.21$ cm.

The radial boundary values $\rho_{\min}(Z)$ and $\rho_{\max}(Z)$ were determined numerically using a uniform grid search over q_2, q_3, q_4, q_5 within [−π,π], with a discretization step of 5°, the resulting range of Z-coordinates was divided into 150 horizontal layers, and the following is obtained: $\rho_{\max} = 94.83$ cm.

Herein, since the q_6 angle does not affect the position of the TCP, it is adopted that $q_6 = 0$. For the adopted numerical values of DH parameters, the inner and outer boundaries of the meridional cross-section are depicted in Fig. 4.

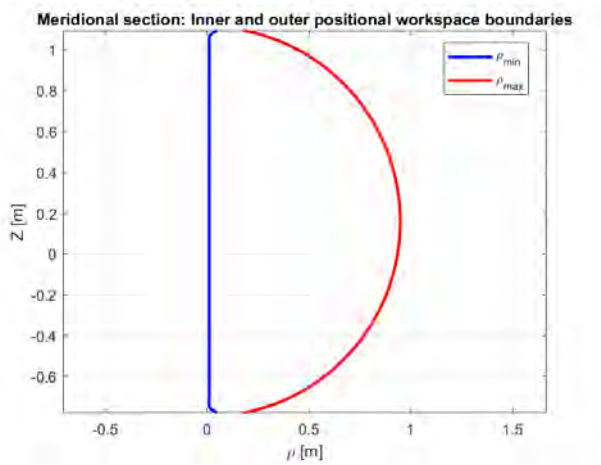


Figure 4 The inner and outer boundaries of the meridional cross-section.

The workspace envelope obtained using Approach 1 is shown in Fig. 5. By rotating $\rho_{\min}(Z)$ and $\rho_{\max}(Z)$ around the z_0 axis, the positional workspace envelope for the ARma6 robot is obtained from 120 angular samples.

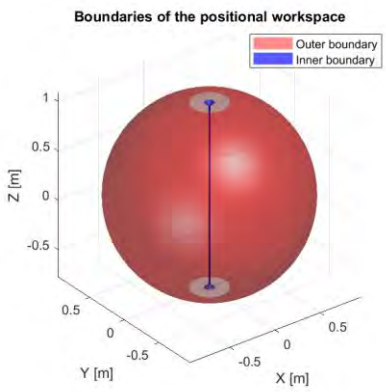


Figure 5 Positional workspace envelope for Arma6 robot obtained by applying Approach 1

To compare the two approaches, Fig. 6 superimposes 25 circular TCP paths corresponding to the effective reach $R_{1\max}$, onto the numerically determined workspace envelope by Approach 1. The circular paths closely follow the outer boundary obtained using Approach 1, indicating good agreement between the two approaches.

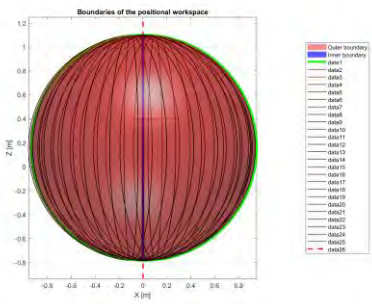


Figure 6 Positional workspace envelope for ARma6 robot obtained by applying Approach 1 and Approach 2 at 25 positions of the $R_{1\max}$ circle around the BASE FRAME vertical axis.

4. DISCUSSION

The proposed methodology combines analytical and numerical approaches for evaluating the kinematic characteristics of articulated manipulators. While the forward kinematic model provides analytical expressions for the TCP position, numerical optimization enables an accurate determination of the maximum reach without relying on exhaustive discretization of the joint space. Furthermore, exploiting the rotational symmetry of the manipulator positional workspace allows the analysis to be performed for a single value of the first joint angle, significantly reducing the computational effort required for three-dimensional workspace determination.

The obtained reach and position workspace envelope provide useful geometric information during the conceptual design stage, i.e. quantitative measures and quick assessment of the robot operating range and the comparison of different manipulator geometries. Since both characteristics depend directly on the selected DH parameters, the proposed methodology can support the selection of appropriate link dimensions in the early stages of robot design before detailed structural and dynamic analyses are performed. When designing a robot, an increase in reach is obtained by increasing the length of the robot's links, however this causes an increase in the mass of the links, the required motor torques, a decrease in stiffness and potentially the accuracy, and the robot may be more prone to vibrations.

In future work, the obtained reach and position workspace envelope can be incorporated into multi-objective optimization procedures aimed at identifying the optimal geometric parameters of the robot. Such optimization could simultaneously consider workspace characteristics, reach, structural stiffness, payload capacity, joint torques, and manipulator mass, resulting in improved overall robot performance. Additionally, the presented methodology should be extended by incorporating self-collision constraints.

5. CONCLUSION

This paper presented the determination of the reach and the position workspace envelope for the six DoF articulated robot ARma6 using its analytical FK model. The reach was obtained numerically using Monte Carlo sampling and nonlinear optimization, and the limits of Cartesian coordinates were determined analytically using trigonometric identities. Two approaches were applied to the position workspace envelope determination. In both approaches, the analysis was performed for a constant value of the first joint angle, and the resulting two-dimensional geometry was subsequently revolved about the base frame vertical axis in order to obtain the three-dimensional workspace envelope. In Approach 1, the rotational symmetry of the manipulator positional workspace was exploited to reconstruct the workspace envelope from the numerically determined inner and outer boundaries of a single meridional cross-section. In Approach 2, the outer workspace envelope was determined geometrically from the circular TCP path

corresponding to the effective robot reach. The obtained numerical results characterize the reach and position workspace of the current ARma6 geometry and provide a basis for further evaluation of the manipulator geometry during the design process. Future work will extend the workspace analysis by incorporating self-collision constraints and investigating the influence of the robot geometric parameters on the usable workspace using a multi-objective optimization framework. Additionally, the obtained workspace envelope will be integrated and visualized within the ARma6 XR environment to support intuitive robot task planning.

Acknowledgment

This research has been supported by the research grants of the Serbian Ministry of Science, Technological Development, and Innovations, Grant No. 451-03-33/2026-03/ 200066

References

[1] ONG, S.K., YEW, A.W.W., THANIGAIVEL, N.K., NEE, A.Y.C.: Augmented Reality-Assisted Robot Programming System for Industrial Applications, Robot. Comput.-Integr. Manuf., 61 (2020) 101820.

[2] DEVIC, A., VIDAKOVIC, J., ZIVKOVIC, N.: Development of Standalone Extended-Reality-Supported Interactive Industrial Robot Programming System, Machines, 12(7) (2024) 480.

[3] RIVERA-PINTO, A., KILDAL, J.: From Digital Twin Towards Shop Floor: Codeless Programming of Pick-and-Place Robots Through Extended-Reality Interaction, Proceedings of the Conference on Production Systems and Logistics: CPSL 2026, Offenburg, 2026, pp. 391–400.

[4] BROGÅRDH, T.: Present and Future Robot Control Development—An Industrial Perspective, Annu. Rev. Control, 31 (2007) 69–79.

[5] VIDAKOVIC, J., DEVIC, A., DIMIC, Z., ZIVKOVIC, N., LAZAREVIC, I.: A Solution to the Forward Kinematic Problem of the ARma6 Robot, Proceedings of the 25th International Symposium INFOTEH-JAHORINA, 2026.

[6] ASPRAGATHOS, N.A., DIMITROS, J.K.: A Comparative Study of Three Methods for Robot Kinematics, IEEE Trans. Syst., Man, Cybern. B, 28(2) (1998) 135–145.

[7] LYNCH, K.M., PARK, F.C.: Modern Robotics: Mechanics, Planning, and Control, Cambridge University Press, Cambridge, 2017.

[8] SPONG, M.W., VIDYASAGAR, M.: Robot Dynamics and Control, John Wiley & Sons, New York, 2008.

[9] ORLANDO, E.M., NENNA, F., LORUSSO, F.M., MINGARDI, M., BUODO, G., GAMBERINI, L.: Capturing Shared Actions and Coordination in Collaborative Workspaces: A Markerless Pipeline for Synchronized Human–Robot 3D Pose Estimation, IEEE Access, 14 (2026) 67101–67114.

[10] DENAVIT, J., HARTENBERG, R.S.: A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices, J. Appl. Mech., 22 (1955) 215–221.

[11] SICILIANO, B., SCIAVICCO, L., VILLANI, L., ORIOLO, G.: Robotics: Modelling, Planning and Control, Springer, London, 2009.

[12] KVRGIC, V.: Development of Intelligent System for Control and Programming of Industrial Robots, Doctoral Dissertation, Faculty of Mechanical Engineering, University of Belgrade, Belgrade, 1998.

[13] METROPOLIS, N., ULAM, S.: The Monte Carlo Method, Journal of the American Statistical Association, 44(247) (1949) 335–341.

[14] LAVALLE, S.M., BRANICKY, M.S., LINDEMANN, S.R.: On the Relationship Between Classical Grid Search and Probabilistic Roadmaps, International Journal of Robotics Research, 23(7–8) (2004) 673–692.

[15] THE MATHWORKS, INC.: MATLAB, The MathWorks, Inc., Natick, Massachusetts, US

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

STOCHASTIC ERROR CHARACTERIZATION AND KALMAN FILTERING OF A MEMS GYROSCOPE BASED ON ALLAN VARRIANCE ANALYSIS

ADRIJANA JOVANOVIĆ
School of Electrical Engineering, Belgrade,
Military Technical Institute, Belgrade, jovanovicadrijana14@gmail.com

JELENA POPOVIĆ-BOŽOVIĆ
School of Electrical Engineering, Belgrade, jelena@etf.bg.ac.rs

ZORAN KRSTIĆ
Military Technical Institute, Belgrade, trik17@proton.me

Abstract: *Micro-Electro-Mechanical Systems (MEMS) gyroscopes are widely used in modern inertial navigation and attitude determination systems due to their compact size, low power consumption, and high integration capability. However, their performance is limited by stochastic errors that significantly affect long-term measurement accuracy, especially in navigation applications. This paper presents the stochastic error characterization of a MEMS gyroscope from the STIM300 inertial measurement unit using Allan variance analysis. The main noise components, including angle random walk, bias instability, and rate random walk, are identified and quantified from experimental measurement data. The obtained noise parameters are used for the development of a Kalman filter model aimed at reducing the influence of stochastic errors and improving gyroscope output stability. Experimental results demonstrate the effectiveness of the proposed approach in reducing measurement drift and enhancing the reliability of MEMS-based inertial systems.*

Keywords: *MEMS gyroscope, STIM300, Allan variance, stochastic errors, angle random walk, bias instability, Kalman filter, inertial navigation system.*

1. INTRODUCTION

Inertial Measurement Units (IMUs) are fundamental components of modern navigation, guidance, stabilization, and control systems. By combining gyroscopes and accelerometers, they provide continuous measurements of angular velocity and specific force required for estimating the motion and orientation of a vehicle. The overall performance of inertial navigation systems is therefore strongly dependent on the accuracy and stability of inertial sensors, particularly gyroscopes.

Micro-Electro-Mechanical Systems (MEMS) gyroscopes have become widely adopted due to their compact size, low power consumption, high reliability, and compatibility with integrated electronic systems. Their favorable balance between cost and performance has enabled their application in unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs), autonomous platforms, robotics, industrial automation, and tactical-grade navigation systems [1], [2]. Despite these advantages, MEMS gyroscopes remain significantly more sensitive to stochastic error sources than navigation-grade fiber optic and ring laser gyroscopes, resulting in reduced long-term measurement accuracy [1], [3].

The dominant stochastic errors include angle random walk, bias instability, and rate random walk. Errors arise from inherent electronic noise, thermal effects, and other internal sensor imperfections. These noise processes accumulate over time and introduce drift into inertial navigation solutions, making accurate stochastic modeling essential for reliable attitude and navigation estimation. While deterministic errors can be compensated through calibration procedures, stochastic errors require statistical characterization and appropriate estimation techniques.

Among the available methods, Allan variance analysis has become the standard technique for stochastic characterization of inertial sensors and gyroscopes, as specified in IEEE Std 952-1997 [4]. It enables the identification and quantification of the dominant stochastic noise processes affecting inertial sensor performance [3], [5], [6]. By analyzing stationary sensor data over different averaging times, Allan variance enables the estimation of noise coefficients that serve as the basis for stochastic sensor modeling and filter design [3], [6], [7], [8].

This paper investigates the stochastic characteristics of the MEMS gyroscope integrated in the STIM300 inertial measurement unit (IMU) manufactured by Sensonor [9]. The stochastic parameters obtained from Allan variance analysis are incorporated into a Kalman filter model to estimate gyroscope bias and suppress measurement noise. The proposed methodology combines experimental noise characterization with model-based filtering to improve the stability and reliability of MEMS gyroscope measurements.

Corresponding author: A. Jovanovic;
E-mail: jovanovicadrijana14@gmail.com

DOI: 10.5937/OTEH26073J

The paper is organized as follows. Section 2 describes the STIM300 MEMS inertial measurement unit and the experimental setup used for data acquisition. Section 3 presents the stochastic error characterization based on Allan variance analysis, with particular emphasis on Angle Random Walk, Bias Instability, and Rate Random Walk. In Section 4, a state-space model of the gyroscope is developed and a Kalman filter is designed for angular-rate estimation based on the identified stochastic error parameters. Section 5 presents the simulation and experimental results, including the evaluation of the filtering performance using statistical error metrics. Finally, Section 6 summarizes the main findings and conclusions of the study.

2. MEMS GYROSCOPES OPERATING PRINCIPLE AND ERROR SOURCES

2.1. Operating principle of MEMS gyroscopes

MEMS gyroscopes measure angular velocity by exploiting the Coriolis effect acting on a vibrating mechanical structure. A proof mass is continuously driven to oscillate along the drive axis. When the sensor experiences angular motion about the sensitive axis, the Coriolis force induces a secondary vibration perpendicular to the original motion. The amplitude of this vibration is proportional to the applied angular velocity and is converted into an electrical signal using capacitive, piezoresistive, or piezoelectric sensing elements.
The Coriolis force can be expressed as:

$$F_c = -2m(\vec{w} \times \vec{v}), \qquad (1)$$

where F_c represents the Coriolis force, m is the proof mass, $\vec{w}$ is the angular velocity vector, and $\vec{v}$ is the velocity of the vibrating mass.

The generated electrical signal is subsequently conditioned, compensated, and converted into the measured angular rate. The sensitivity and stability of the sensor depend on the mechanical characteristics of the vibrating structure, including its resonant frequency, quality factor, and damping. Advances in MEMS fabrication technology have significantly improved these characteristics, enabling devices such as the STIM300 IMU to achieve navigation-grade performance while maintaining compact dimensions and low power consumption.

2.2 Error sources in MEMS gyroscopes

The output of a MEMS gyroscope contains both deterministic and stochastic error components and can be generally expressed as:

$$w_m(t) = w(t) + \delta w_d(t) + \delta w_s(t)\,, \qquad (2)$$

where $w_m(t)$ represents the measured angular rate, $w(t)$ is the true angular rate, $\delta w_d(t)$ represents deterministic errors, and $\delta w_s(t)$ represents stochastic errors.

Deterministic errors, such as bias offset, scale factor error, axis misalignment, and temperature-dependent effects, are repeatable and can be significantly reduced through calibration procedures. Since the objective of this work is the characterization and filtering of stochastic errors, deterministic error sources are not considered further.

The stochastic component represents the dominant limitation of long-term gyroscope performance because it varies randomly with time and cannot be removed through calibration. Consequently, an appropriate stochastic model is required for accurate state estimation and filtering.

The stochastic component of the gyro output from expression (2) can be simplified as:

$$w_m(t) = w(t) + b(t) + n(t), \qquad (3)$$

where $w(t)$ is the true angular rate, $b(t)$ is the gyro bias, and $n(t)$ represents the measurement noise.

3. ALLAN VARIANCE – BASED NOISE CHARACTERIZATION

Allan variance is a widely used method for characterizing stochastic errors in inertial sensors. Unlike conventional variance analysis, Allan variance enables the identification of different noise processes by evaluating sensor stability over a range of averaging times. Dynamic extensions of Allan variance have also been proposed to analyze non-stationary stochastic processes and time-varying sensor characteristics [7], [10], [11].

For a discrete sequence of gyro measurements, the Allan variance is defined as:

$$\sigma_A^2(\tau) = \frac{1}{2(N-1)}\sum_{i=1}^{N-1}[\overline{w}_{k+1}(\tau) - \overline{w}_k(\tau)]^2, \qquad (4)$$

where τ is the averaging time, N represents number of non--overlapping averaging intervals and $\overline{w}_k(\tau)$ represents the average angular rate over the interval τ.

The Allan deviation is obtained as:

$$\sigma_A(\tau) = \sqrt{\sigma_A^2(\tau)}\,. \qquad (5)$$

The Allan deviation is usually presented on a logarithmic scale as a function of the averaging time. Different stochastic error sources produce characteristic slopes in the Allan deviation curve, as shown in Figure 1, enabling their identification.

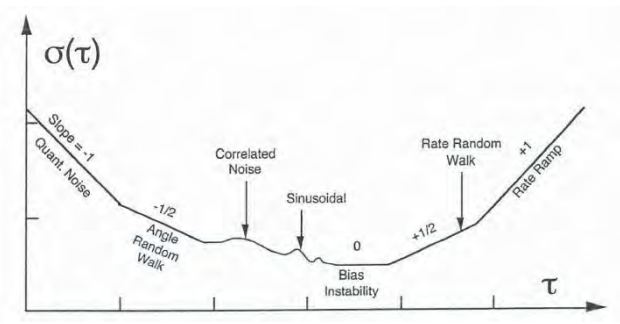


Figure 1 – Sample plot of Allan variance analysis results (IEEE 952 1997) [4]

The main stochastic error components considered in this work are:

3.1 Angle Random Walk

Angular Random Walk represents the high-frequency white noise component of the gyroscope output. In the Allan deviation plot, ARW is identified from the region where the curve has a slope of −1/2 logarithmic coordinates.

The Allan deviation in this region can be expressed as [4]:

$$ARW = \sigma_A(\tau)\sqrt{\tau}\,, \tag{6}$$

where ARW represents the angle random walk coefficient. Then,

$$\sigma_{ARW} = \frac{ARW}{\sqrt{T_s}}\,, \tag{7}$$

where σ_{ARW} is standard deviation of the discrete white Gaussian nosie used for syntetic signal generation and T_s is sampling period. The corresponding noise is modeled as a white Gaussian noise process:

$$n_{ARW} \sim N(0, \sigma_{ARW}^2). \tag{8}$$

3.2 Bias instability

Bias instability represents slow variations of the gyro bias and determines the long-term stability of the sensor. It is caused by low-frequency noise processes, including flicker noise. It appears as the minimum point of the Allan deviation curve. The bias instability coefficient is obtained from the corresponding Allan deviation value at the zero-slope region.

The corresponding Allan deviation value represents the bias instability coefficient:

$$BI = \frac{\sigma_{A,min}}{\sqrt{\frac{2\,ln(2)}{\pi}}}\,. \tag{9}$$

The averaging time at which this minimum occurs represents the characteristic time scale of the bias variation.

Unlike ARW, Bias Instability cannot be modeled as a white noise process. Therefore, the bias drift of a MEMS gyroscope is modeled as a first-order Gauss–Markov stochastic process. The continuous-time bias model is described by:

$$\dot{b}(t) = -\frac{1}{\tau_b}b(t) + w_b(t)\,, \tag{10}$$

where $b(t)$ represents the gyroscope bias, τ_b is correlation time and $w_b(t)$ is a zeromean white noise process.

The discrete-time form of the Gauss–Markov model is obtained as

$$b_{k+1} = e^{-\frac{\Delta t}{\tau_b}}b_k + w_k\,. \tag{11}$$

Here, Δt represents the sampling interval and τ_b defines the temporal correlation of the bias process. A larger value of τ_b corresponds to slower bias variations and stronger correlation between consecutive samples.

The covariance of the bias driving noise is:

$$Q_{BI} = BI^2(1-\phi^2), \tag{12}$$

where $\phi = e^{-\Delta t/\tau_b}$ and τ_b is the correlation time of the bias process.

The obtained BI parameter defines the magnitude of bias uncertainty, while the correlation time determines the temporal behavior of the bias drift.

3.3 Rate Random Walk

Rate Random Walk describes the long-term accumulation of angular rate errors. It appears as a slope of +1/2 on the Allan deviation curve. The corresponding coefficient can be estimated from:

$$RRW = \sigma_A(\tau)\sqrt{\frac{3}{\tau}}\,, \tag{13}$$

where RRW represents the rate random walk coefficient.

In the Kalman filter, RRW is modeled as a random walk process and contributes to the process noise covariance:

$$Q_{RRW} = RRW^2\Delta t. \tag{14}$$

4. STOCHASTIC ERROR MODEL PARAMETER ESTIMATION AND KALMAN FILTER DESIGN

The stochastic parameters obtained from Allan variance analysis provide a quantitative description of the gyro noise characteristics. These parameters are used to define the stochastic error model and determine the covariance matrices required for Kalman filter implementation [3], [12], [13]. The relationship between Allan variance coefficients and Kalman filter parameters has been extensively investigated and provides a practical methodology for selecting appropriate process and measurement noise covariance matrices [12].

In this work, the gyro bias is modeled as an additional state variable and estimated together with the angular velocity using a discrete Kalman filter. Such an approach enables simultaneous bias estimation and measurement noise suppression, resulting in improved long-term stability of the estimated angular rate [1], [2], [14].

The discrete state-space model can be expressed as:

$$x_{k+1} = Fx_k + Gw_k\,, \tag{15}$$

$$z_k = Hx_k + v_k\,, \tag{16}$$

where x_{k+1} is the state vector, F is the state transition matrix, w_k represents the process noise, G represents noise

gain matrix, z_k is the measurement vector, H is the measurement matrix, v_k represents the measurement noise.

The state vector is defined as:

$$x_k = \begin{bmatrix} w_k \\ b_k \end{bmatrix}, \tag{17}$$

where w_k represents the estimated angular rate and b_k represents the gyro bias.

The state transition matrix describes the dynamic behavior of the system between two consecutive sampling instants. The angular velocity is considered constant during one sampling interval, while the bias follows a first-order Gauss-Markov process.

Therefore, the state transition matrix is defined as:

$$F_k = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix}, \tag{18}$$

where $\phi = e^{-\Delta t/\tau}$ is the correlation coefficient of the Gauss-Markov bias model. The parameter τ represents the correlation time of the bias. It determines how quickly the bias loses correlation with its previous value. The correlation time τ was selected according to the bias instability averaging time specified in the sensor datasheet.

The process noise covariance matrix is defined as:

$$Q = \begin{bmatrix} RRW^2 \Delta t & 0 \\ 0 & BI^2(1-\phi^2) \end{bmatrix}, \tag{19}$$

where RRW is the Rate Random Walk coefficient obtained from Allan variance analysis, BI is the Bias Instability coefficient, Δt is the sampling interval and ϕ is the Gauss-Markov correlation coefficient.

The first diagonal element represents the uncertainty of the angular velocity state caused by long-term rate variations. The second diagonal element represents the uncertainty of the bias state caused by the stochastic variation of the MEMS gyroscope bias.

The appropriate selection of the matrix Q is essential for Kalman filter performance. If the process noise is underestimated, the filter becomes insensitive to actual sensor variations, while excessive process noise leads to increased estimation uncertainty.

The measurement matrix defines the relationship between the system states and the measured gyroscope output.

The measurement equation is:

$$z_k = w_k + b_k + n_k. \tag{20}$$

Therefore, the measurement matrix is:

$$H_k = [\,1 \quad 1\,]. \tag{21}$$

This means that the measured angular velocity contains contributions from both the true angular velocity and the sensor bias.

The measurement noise covariance matrix describes the uncertainty of the gyroscope measurement.

The main short-term noise component of MEMS gyroscopes is Angle Random Walk (ARW), which appears as white noise in the angular rate measurement.

The measurement noise covariance matrix is defined as:

$$R_k = \frac{ARW^2}{\Delta t}, \tag{22}$$

where ARW represents the Angle Random Walk coefficient obtained from Allan variance analysis.

The R matrix determines how much confidence the Kalman filter assigns to the raw gyroscope measurement.

A larger value of R means that the measurement is considered less reliable, while a smaller value indicates higher confidence in the sensor output.

During the prediction stage, the previous state estimate is propagated using the state transition matrix:

$$\hat{x}_{\bar{k}|k-1} = F_k \hat{x}_{k-1|k-1}\,. \tag{23}$$

The predicted covariance matrix is calculated as:

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k\,, \tag{24}$$

where P_k is the state covariance matrix and Q_k represents process uncertainty.

The Kalman gain is calculated using:

$$K_k = P_{k|k-1} H_k^T \left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1}. \tag{25}$$

The state estimate is corrected according to:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1})\,. \tag{26}$$

The covariance matrix is updated as:

$$P_{k|k} = (I - K_K H_k) P_{k|k-1}\,. \tag{27}$$

5.EXPERIMENTAL SETUP AND RESULTS

The STIM300 inertial measurement unit, shown in Figure 2, was mounted on a laboratory table and remained stationary throughout the experiment to eliminate external rotational motion. Gyroscope data were collected for approximately three hours at a sampling frequency of 500 Hz using the manufacturer's communication interface. The recorded measurements were exported to MATLAB, where Allan variance analysis and Kalman filtering were performed offline. Stationary measurements were selected because they allow accurate identification of stochastic sensor errors without contamination by vehicle dynamics. As the gyroscope remained stationary with respect to the laboratory frame, the contribution of the Earth's rotation to the measured angular rate was neglected, and the true

angular rate was assumed to be zero for the purposes of stochastic error characterization and Kalman filter evaluation.

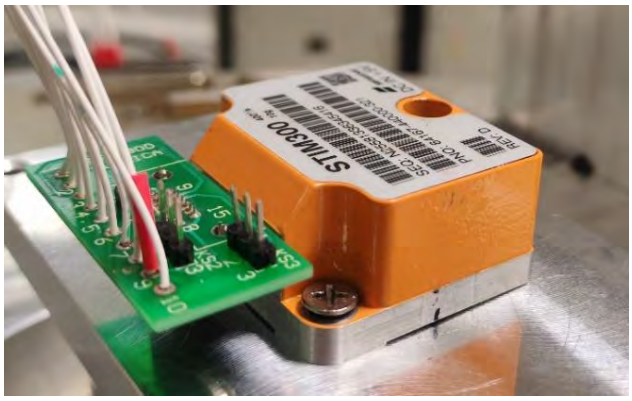


Figure 2. IMU STIM300

The Allan deviation curve, shown in Figure 3, clearly exhibits three characteristic regions corresponding to Angle Random Walk, Bias Instability and Rate Random Walk.

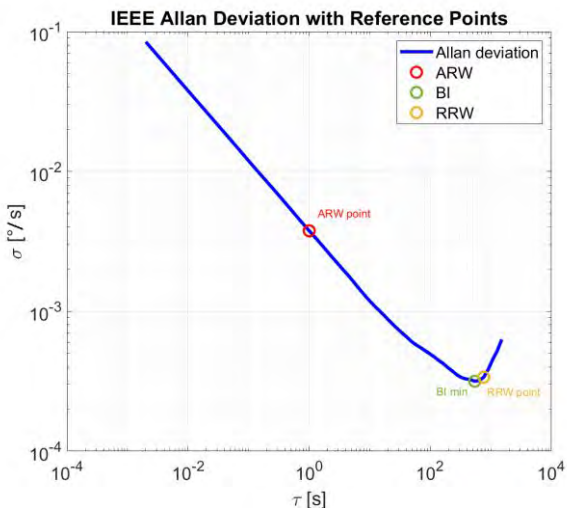


Figure 3. Allan Deviation

Their estimated coefficients, together with the corresponding datasheet values, are summarized in Table 1 and subsequently used for stochastic model identification.

Table 1. Value of stochastic error components

Noise parameter	Symbol	Value	Sensor datasheet specifications	Unit
Angle Random Walk	ARW	0.0038	0.0025	$°/\sqrt{s}$
Bias Instability	BI	0.0004734	0.0000833	$°/s$
Rate Random Walk	RRW	0.000021239	/	$°/s^{3/2}$

5.1 Simulation Results

A synthetic gyro signal was generated based on the stochastic error parameters identified from the Allan deviation analysis. The true angular rate was assumed to be zero, corresponding to the stationary condition of the sensor. The Angle Random Walk component was modeled as white Gaussian noise, with its standard deviation determined from the identified ARW coefficient. The Rate Random Walk component was generated as integrated white Gaussian noise, representing the slowly varying stochastic component of the angular-rate measurement. The Bias Instability component was modeled as a first-order Gauss–Markov process, whose parameters were selected according to the identified bias characteristics. The individual stochastic components were then combined to obtain the synthetic gyro output. In this way, the simulated signal reproduced the main stochastic characteristics identified from the experimental gyro data. The generated synthetic signal was subsequently used as the input to the proposed Kalman filter, allowing the performance of the stochastic model and the filtering algorithm to be evaluated under controlled conditions before their application to the experimental measurements. The corresponding filtering results are presented in Figure 4.

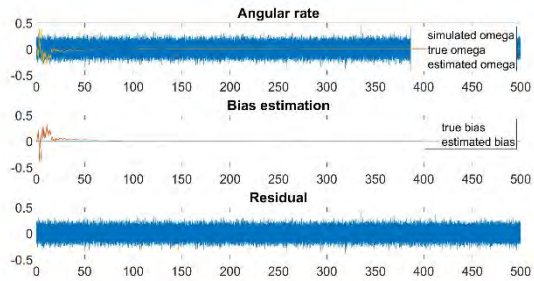


Figure 4. Kalman filter estimation results: simulated and estimated angular rate, true and estimated bias, and residual signal

5.2 Experimental Results

After validating the Kalman filter using simulated data, the same filter configuration was applied to measurements acquired from the STIM300 MEMS gyroscope. During the experiment, the sensor remained stationary in order to isolate stochastic error sources and eliminate the influence of external motion. The gyroscope data were sampled at a frequency of 500 Hz and processed offline in MATLAB.

Unlike the simulation, where the stochastic model is known, the experimental measurements represent the actual behavior of the sensor, including all dominant stochastic error components. The same Kalman filter configuration was subsequently applied to experimentally acquired STIM300 measurements. Since all stochastic parameters were obtained directly from Allan variance analysis, no empirical tuning of the covariance matrices was required. The filtered output exhibits significantly reduced short-term fluctuations while preserving the slowly varying gyro behavior. In addition, the estimated bias follows the expected low-frequency variations

predicted by the Gauss-Markov model.

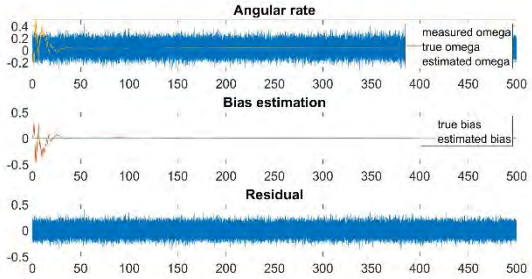


Figure 5. Kalman filter estimation results: measured and estimated angular rate, true and estimated bias, and residual signal

Figure 5 presents the measured and estimated angular rate, the true and estimated bias, and the residual signal. Since the gyroscope remained stationary throughout the measurement, the true angular rate was equal to zero. Consequently, the measured signal consisted only of stochastic errors, primarily white noise (ARW) and slowly varying bias drift.

The estimated angular rate converges rapidly toward zero, indicating that the Kalman filter successfully suppresses measurement noise while preserving the expected system dynamics. The estimated signal exhibits significantly lower variance than the raw sensor measurements, demonstrating the effectiveness of the proposed filtering approach.

The bias estimation initially exhibits relatively large oscillations. This transient behavior is expected because the filter starts from an initial estimate that differs from the actual bias. During the first several seconds, the Kalman filter gradually updates the state estimate as new measurements become available. As the estimation covariance decreases, the bias estimate converges toward the true bias generated by the Gauss–Markov model.

After the transient period, the estimated bias closely follows the true bias, indicating that the adopted stochastic model accurately represents the bias dynamics.

The innovation sequence remains centered around zero without any noticeable long-term trend, suggesting that the selected stochastic model adequately describes the measurement process and that no significant unmodeled dynamics remain.

Overall, the obtained results confirm that the Allan variance parameters, together with the Gauss–Markov bias model, provide an appropriate stochastic representation of the MEMS gyroscope and enable accurate estimation of both the angular rate and the sensor bias.

The performance of the proposed Kalman filter was evaluated using the standard deviation and root mean square error (RMSE), as summarized in Table 2. The standard deviation quantifies the achieved noise reduction, whereas RMSE evaluates the overall estimation accuracy.

Table 2. Performance indicators

Parameter	Value	Unit
Measured standard deviation	0.084858	deg/s
Estimated standard deviation	0.012279	deg/s
Noise reduction	85.46	%
Measured RMSE	0.090061	deg/s
Estimated RMSE	0.034220 deg/s	deg/s
RMSE improvement	62.00	%
Bias RMSE	0.01221466	deg/s

To quantitatively evaluate the filtering performance, the standard deviation and RMSE were calculated for both the measured and Kalman-filtered angular-rate signals. The measured standard deviation was used to quantify the dispersion of the raw gyroscope measurements, while the measured RMSE was used to quantify their error with respect to the assumed true angular rate. Similarly, the estimated standard deviation characterizes the dispersion of the Kalman-filtered signal, while the estimated RMSE quantifies its error with respect to the assumed true angular rate. The noise reduction was calculated from the reduction in standard deviation between the measured and estimated signals, whereas RMSE improvement represents the relative reduction in RMSE after filtering. The bias RMSE was calculated separately to quantify the error of the estimated gyroscope bias with respect to the reference bias. The standard deviation of the measured signal was 0.084858 deg/s, while the estimated signal exhibited a reduced standard deviation of 0.012279 deg/s, corresponding to an 85.46% reduction in noise level. The RMSE decreased from 0.090061 deg/s for the raw measurements to 0.034220 deg/s after filtering, representing a 62% improvement in overall estimation accuracy. The bias estimation error remained low, with a bias RMSE of 0.0122 deg/s, indicating that the Kalman filter successfully tracked the slowly varying gyro bias component.

6. CONCLUSION

This paper presented a methodology for stochastic modeling and filtering of a MEMS gyroscope using experimentally identified noise characteristics. The proposed approach combines Allan variance analysis with a Kalman filter whose covariance matrices are determined directly from sensor-specific stochastic parameters rather than empirical tuning.

The obtained results demonstrate that the proposed filter effectively reduces short-term measurement fluctuations while providing reliable estimation of the slowly varying gyroscope bias.

By linking experimental noise characterization with model-based state estimation, the proposed methodology

provides a practical framework for improving the quality of MEMS gyroscope measurements. Such an approach is particularly suitable for inertial navigation and attitude estimation systems, where accurate stochastic sensor modeling is essential for achieving stable long-term performance.

Future work will focus on extending the proposed methodology to a complete error-state Kalman filter for integrated inertial navigation systems and evaluating its performance under dynamic operating conditions.

References

[1] FARRELL,J.A., SILVA,F.O., RAHMAN,F., WENDEL,J.: IMU *Error Modeling Tutorial: INS State Estimation with Real-Time Sensor Calibration,* IEEE Aerospace and Electronic Systems Magazine**,** 36(2) (2021) 64-79.

[2] NAVIDI,N., LANDRY,R. Jr.*: A New Perspective on Low-Cost MEMS-Based AHRS Determination,* Sensors, 21(4**)** (2021) 1383.

[3] EL-SHEIMY,N., HOU,H.*, NIU,X.: Analysis and Modeling of Inertial Sensors Using Allan Variance,* IEEE Transactions on Instrumentation and Measurement, 57(**1)** (2008) 140-149.

[4] IEEE: *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros*, IEEE Std 952-1997, IEEE, 1998.

[5] ASLAM,Q.H., IQBAL,A., BIBI,F., BIBI,S.: *Error Estimation of Inertial Sensors Using Allan Variance,* IFAC Proceedings Volumes, 41(1) (2008) 167-171.

[6] LV,P., LIU,J., LAI,J., HUANG,K.: *Allan Variance Method for Gyro Noise Analysis Using Weighted Least Square Algorithm,* Optik*,* 126(20) (2015) 2529-2534.

[7] SONG,R., CHEN,X., HUANG,H.: *Nonstationary Dynamic Stochastic Error Analysis of Fiber Optic Gyroscope Based on Optimized Allan Variance,* Sensors and Actuators A: Physical, 276 (2018) 26-33.

[8] STANČIĆ,R. *Razvoj integrisanog sistema inercijalne i satelitske navigacije*, Elektrotehnički fakultet, Beograd, 2010.

[9] SENSONOR: *STIM300 Inertial Measurement Unit Datasheet*, Sensonor AS, TS1524, Rev. 31, February 2024

[10] WANG,L., ZHANG,C., GAO,S., WANG,T., LIN,T., LI,X.: *The Fast Algorithm of the Dynamic Allan Variance for Gyroscopes,* Optik, 127(5) (2016) 2413-2418.

[11] VUKMIRICA, V., TRAJKOVSKI, I., ASANOVIĆ, N., *Two Methods for the Determination of Inertial Sensor Parameters,* Scientific Technical Review, 2010, Vol.60, No.3-4, pp.27-33

[12] VAN DIERENDONCK,A.J., MCGRAW,J.B.: *Relationship Between Allan Variances and Kalman Filter Parameters,* Proceedings of the 16th Annual Precise Time and Time Interval Applications and Planning Meeting*, (1984) 273-293*

[13] XIONG,B.F., TANG,C., WANG,Q., ZHU,S.P.: *Research on the Modeling of Random Drift Error and Filtering Technology of Low Cost MEMS Gyroscope,* Modelling, Measurement and Control A, 90(1**)** (2017) 16-31.

[14] LI,W., WANG,J.: *Effective Adaptive Kalman Filter for MEMS-IMU/Magnetometers Integrated Attitude and Heading Reference Systems,* The Journal of Navigation, 66(1**)** (2013) 99-113.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A LAYERED COUNTER-UAS CONCEPT: FROM MULTI-SENSOR DETECTION AND CLASSIFICATION TO KILL-WEB INTEGRATION

ILIJA POPADIĆ, MILJAN VUČETIĆ, ZORAN BANJAC, BOŠKO BOŽILOVIĆ, IVAN VULIĆ
Vlatacom Institute of High Technologies, Belgrade, ivan.vulic@vlatacom.com

Abstract: *The proliferation of small unmanned aircraft systems (UAS), from commercial quadcopters to Shahed-class loitering munitions, has created a threat that legacy air defense was not designed to counter: low, slow, small, inexpensive, numerous and increasingly autonomous targets. This paper presents a layered counter-UAS (C-UAS) concept spanning the full engagement chain, grounded in engineering results and operational experience. We argue that no single sensing modality suffices: the sensor set must include a passive layer, with acoustic sensing — non-jammable and independent of line of sight — addressing the autonomous, fiber-guided and otherwise radio-silent platforms that defeat radio-frequency (RF) sensing. We then describe artificial-intelligence (AI) classification at the sensor edge, multi-sensor track fusion, and the integration of soft-kill and hard-kill effectors — operational directional jamming and, under development, the AURORA programmable airburst munition, a reusable kinetic micro-interceptor drone and a high-energy laser — into a kill web built on a C7ISTAR solution (command, control, communications, computers, cyber, combat systems and collaboration, with intelligence, surveillance, target acquisition and reconnaissance), whose ORAO (Observe, React, Analyze, Operate) command platform and ARES (Algorithmic Reconnaissance and Execution System) AI core have protected national critical infrastructure since 2024. The concept applies to counter-drone protection of fixed critical-infrastructure sites, vehicles, ships, and temporary sites, including operation in civilian environments.*

Keywords: *counter-UAS, layered defense, multi-sensor fusion, interceptor UAV, kill web.*

1. INTRODUCTION

Few threat classes have evolved as rapidly over the past decade as small unmanned aircraft systems (UAS). Platforms derived from commercial technology are cheap, numerous, simple to operate, and increasingly capable, and loitering munitions of the Shahed type extend the same logic to ranges once reserved for far more expensive weapons. Employed en masse, they impose an unfavorable cost exchange on the defender: over 14,700 one-way attack drones were launched against Ukraine between September 2022 and December 2024, at unit costs one to two orders of magnitude below those of the air-defense missiles initially expended against them [1]. On-board autonomy compounds the problem — a UAS navigating by waypoints or on-board perception offers no control link to detect or deny — and fiber-optically guided first-person-view (FPV) platforms are immune to jamming by construction [2]. It is now broadly recognized in the literature [3], [4], [5] and in strategy and doctrine [2], [6] that no single sensor and no single effector constitutes a counter-UAS (C-UAS) capability. Required is a layered system: complementary sensor modalities, machine-assisted classification, and a command-and-control (C2) layer connecting any sensor to the most appropriate effector — a kill web rather than a linear kill chain. Its measure, and the thread of this paper, is the path from sensor to shooter, which every layer must shorten without surrendering the operator's authority.

C-UAS is a demanding instance of the general command-support problem, whose vocabulary has grown from C2 through C4ISR (command, control, communications, computers, intelligence, surveillance and reconnaissance) to C7ISTAR — command, control, communications, computers, cyber, combat systems and collaboration, integrated with intelligence, surveillance, target acquisition and reconnaissance — which admits cyber defense, combat systems and collaboration as first-class capabilities [7]. The concept is grounded in an integrated anti-drone system — the ORAO (Observe, React, Analyze, Operate) C2 platform carrying the ARES (Algorithmic Reconnaissance and Execution System) artificial-intelligence (AI) core — that has protected a major critical-infrastructure facility in Serbia since 2024, and in research on electro-optical imaging [8], [9], target tracking [10], [11], radar [12], [13] and secure communications. Sections 2 to 6 treat detection, the sensor layer, classification and fusion, decision and effect, and deployment; Section 7 concludes.

2. THE DETECTION PROBLEM

The difficulty begins with physics. Small UAS present radar cross-sections (RCS) orders of magnitude below those of manned aircraft and fly where the radar picture is dominated by clutter, birds and multipath; their speeds reach down to hover, inside the rejection region of classical moving-target processing; thermal signatures are modest, acoustic output attenuates quickly, and visual size defeats unaided observation beyond short range. An emitting UAS offers a radio-frequency (RF) signature buried in congested spectrum; a non-emitting one offers RF sensing nothing — decisive for Shahed-class and other autonomous threats.

Corresponding author: I. Vulić;
E-mail: ivan.vulic@vlatacom.com
DOI: 10.5937/OTEH26074P

Detection ranges are short, closing times shorter, and whatever the architecture achieves must be achieved at a manageable false-alarm rate, because a system that alerts on every bird will be switched off. Experience with a thermal panoramic sensor is instructive [9], [14]: a head covering the full circle in 2.3 s produces some fifty raw detections per second, dominated by clouds and birds; tracking, cloud rejection, bird filtering and AI classification of close targets reduce that flow to about one plausible drone detection per second. Detection is cheap and noisy; tracking, classification and fusion of complementary modalities produce a usable picture [3], [15].

3. THE MULTI-SENSOR DETECTION LAYER

Radar is the backbone of volumetric search: all-weather, day and night, with direct measurement of range and radial velocity; against small UAS, its discriminative asset is the micro-Doppler signature of rotating rotors [15], [16]. In the described system, a compact X-band pulse-Doppler radar is integrated as a third-party component, detecting targets with an RCS of about 0.01 m² at a few kilometers; earlier radar research [12], [13], [17] informs its integration and interpretation of small-target returns. The electro-optical layer compensates for its weak elevation scanning.

RF sensing is passive and continuous, and two of its functions must be distinguished. A wideband receiver performing real-time spectrum analysis over roughly 20 MHz to 6 GHz detects and classifies control, telemetry, and video links, but spectrum analysis alone yields no direction. A direction-finding (DF) antenna array measures bearing by comparing amplitudes across directional elements or using phase interferometry; bearings from two or more DF sensors are triangulated into an approximate position of the platform and, when its controller transmits, of the operator [4]. The limitation is categorical: RF sensing cannot detect non-emitting platforms, such as fiber-guided drones or autonomous ones in radio silence, and must work in dense civilian spectrum.

Electro-optical and infrared (EO/IR) sensing provides what no other modality can: an image — identification, payload assessment, and the visual confirmation that engagement decisions in most settings legally require. The VMSIS3 multi-sensor imaging turret fills this role with a cooled mid-wave (MWIR) or uncooled long-wave infrared (LWIR) channel, a high-definition day channel with optical zoom, a low-light channel and a laser rangefinder from which target coordinates are computed [8], [9]. On an extremely agile pan-tilt positioner, the ROBIN configuration slews to a cued target, and its on-board AI verifies a drone, aircraft, helicopter, or bird in under two seconds. EO/IR is narrow-field and visibility-dependent: most effective when cued rather than searching.

Acoustic sensing detects the propulsion signature of the target: fully passive, non-jammable, effective without line of sight and against precisely the radio-silent platforms — autonomous or fiber-guided — that defeat the RF layer, with 360° coverage at low unit cost and power; propagation modeling draws on earlier acoustic work [18]. Recent peer-reviewed results confirm the modality's maturation [19], and doctrine now recommends passive RF and acoustic layers alongside active sensors [2]. Its physics defines its role — short-to-medium-range passive cueing and gap-filling — and it earns its place exactly where the other modalities are weakest. Table 1 summarizes the complementarity of the four modalities.

Table 1. Complementarity of sensor modalities in the C-UAS detection layer

Modality	Principal strengths	Principal limitations	Role in the layered system
Radar (X-band)	Volumetric all-weather search; range and radial velocity; micro-Doppler	Ground clutter; low-velocity regime; weak elevation scanning	Backbone of early warning and tracking
RF sensing	Passive; link classification; DF bearings and triangulation	Blind to non-emitting targets; congested spectrum	Wide-area passive tripwire and attribution
EO/IR (VMSIS3)	Identification and evidence; AI verification in < 2 s; laser ranging	Weather and visibility dependent; narrow field; requires cueing	Confirmation, identification and geo-location on cue
Acoustic	Passive, non-jammable, 360°; RF-silent UAS without line of sight	Short range; sensitive to wind and ambient noise	Persistent passive detection of azimuth, elevation and velocity

4. CLASSIFICATION AND TRACK FUSION

4.1. Separating targets from confusers

The first task of the intelligence layer is binary and unforgiving: is this track a UAS, or one of the innumerable confusers of the low-altitude environment — above all birds and clouds? In the described system, it is performed by ARES, the Algorithmic Reconnaissance and Execution System: the modular AI core of the C7ISTAR solution, which turns sensor data into a common operational picture and is organized in three tiers that follow the levels of command — ARES Edge (tactical) at the sensor, ARES Core (operational) for fusion across sites, and ARES Strategic (command) for language-model decision support. ARES Edge runs inference at the sensor itself, on a graphics processing unit (GPU) module delivering more than one hundred trillion operations per second, with deep-learning models distinguishing drone, bird, aircraft, and helicopter; only the track and its class label then traverse the network, which keeps the system usable on constrained links and limits data exposure, a concern formalized for distributed sensor networks in [20], [21]. With processing latency below 100 ms, it marks the target, assigns class and confidence, and drives automatic tracking; a Kalman filter with deep-learning re-identification holds the track through occlusion and maneuver, building on robust Kalman tracking in short-wave infrared (SWIR) imagery [10] and motion-triggered track activation [11]. Acoustic classification complements imagery when the target is beyond visual range or emits no visual signature.

4.2. Threat level, fusion and the model lifecycle

Beyond the binary decision, classification refines platform class, link type, payload indications, and signature type, and threat-level estimation combines what the object is with what it is doing — trajectory, altitude, speed, coordination — into a ranked picture. Fusion associates detections across modalities into single system tracks carrying classification, confidence, and full sensor provenance [22], [23]; gossip-type distributed estimation algorithms address consensus across heterogeneous, asynchronously communicating sensor networks [24]. The purpose is continuity — the acoustic layer holding a target behind cover, the radar holding one lost to glare — and credibility: a track confirmed by independent modalities justifies a response that a single-sensor track does not. Because the domain is adversarial and non-stationary, models are maintained assets — curated datasets, monitored in-service performance, controlled retraining [16] — and local training on the edge platform adapts them to the threat picture of a site without data leaving the security perimeter, a sovereignty requirement.

5. FROM SENSOR TO SHOOTER: DECISION, EFFECT AND KILL-WEB INTEGRATION

5.1. The ORAO command-and-control platform

ORAO (also the Serbian word for eagle) is the layer that turns sensing into effect: a C2 platform built on a scalable microservices architecture that underpins the C7ISTAR solution. It integrates sensors and effectors into a single operational picture, suppresses false alarms through fusion, and lets the operator act remotely; new sensors, effectors, or functions are added without altering the core, so that one architecture spans the configurations of Section 6. Figure 1 shows the platform fusing the bearing of the RF direction-finding sensor with electro-optical tracking of the same target, with jammer and sensor sectors on the map.

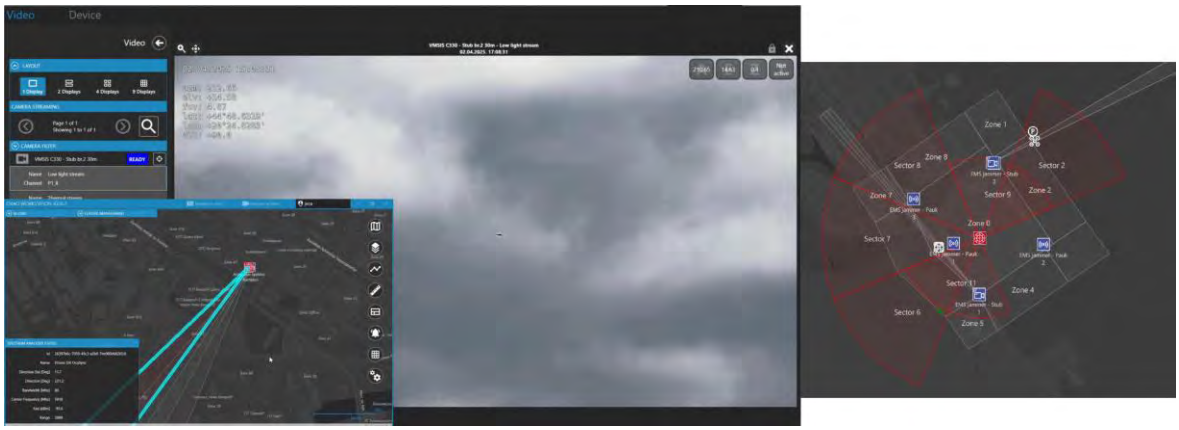


Figure 1. The ORAO console: live optical stream with telemetry, bearing lines of the RF direction-finding sensor with its status panel (classified link, bearing, standard deviation), and sensor and jammer sectors

5.2. The kill web

The integrating principle is the kill web (Figure 2): every sensor and every effector is a node on the common C2 fabric, and the system composes the engagement path best suited to each track — any sensor, best effector [25] — a principle the recent literature refines into federation scoped to concrete engagement chains rather than all-to-all interoperability [26]. The acoustic sensor may hold an RF-silent target the radar cannot separate from clutter; ORAO cues the VMSIS3 turret to that bearing, ARES Edge confirms the object in under two seconds and the rangefinder fixes its coordinates; ARES Core holds the fused track, while ARES Strategic couples the picture to language-model agents that give the operator situation reports and ranked courses of action; on the operator's decision, the appropriate effector is assigned. The fabric is resilient — redundant paths, graceful degradation — and efficient, using the cheapest adequate effector per engagement. It is also a security fabric: every node's identity must be established and its data protected, for which public key infrastructure (PKI) models for authenticating Internet-of-Military-Things devices [27], for interoperability across national and coalition PKI domains [28] and for time-stamped trust in device data [29] provide the design basis. ORAO is supervised-autonomous: AI and rule-based actions prepare and, where configured, execute responses, but authority over the use of force rests with the operator except where narrowly bounded, pre-authorized automatic responses have been deliberately enabled. Where a kill chain offers one path and fails when any link fails, the web offers many, selecting at each engagement the shortest path the rules of engagement allow.

5.3. Soft-kill and hard-kill effectors

Effectors divide into two families with complementary profiles. Soft kill — directional RF jamming of control, telemetry and satellite-navigation links in the standard drone bands — is low-cost per engagement, deep in magazine and comparatively benign; it is the proven method on which the operational system relies, driving the target to descend, return or land. It is ineffective against fiber-guided platforms and autonomous ones navigating without satellite signals, and its use of civilian spectrum is regulated. Hard kill is effective regardless of autonomy and is pursued along three complementary lines.

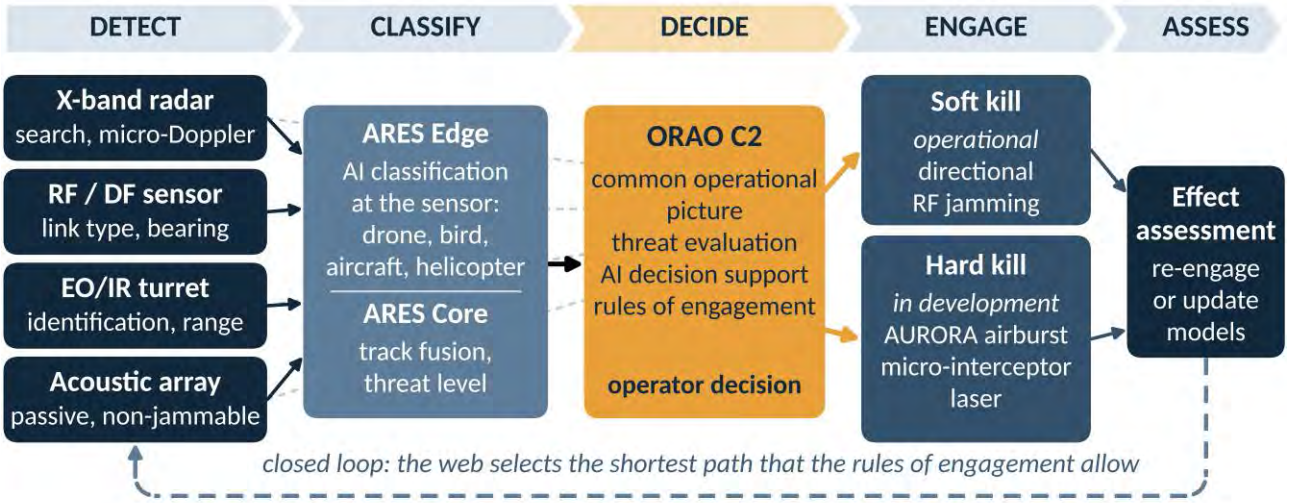


Figure 2. The C-UAS kill web along the engagement sequence: sensors, ARES classification and fusion, ORAO C2 under the rules of engagement, soft kill (operational), hard kill (in development), and effect assessment

Figure 3. Concept rendering of micro-interceptors of the 7-inch class at readiness in a five-position dock-and-launch module; launch is authorized through ORAO over a secure C7ISTAR link

The first line is gun-based: AURORA, a dual-caliber programmable airburst munition for 30 mm and 23 mm autocannons, at technology readiness level 2 with a roadmap to level 7. Instead of requiring a direct hit on a small, agile target, the round carries an electronic fuze built around a high-g-resistant field-programmable gate array (FPGA) timer, set inductively at the muzzle from the muzzle velocity measured for each round; it detonates at the computed intercept point, ejecting a forward cone of pre-formed tungsten fragments that saturates the flight path and defeats the target from a near miss. Timing accuracy is of the order of 2 ms, and the design objective is a kill probability above 85% against agile UAS at a cost per engagement that conventional high-explosive fire cannot approach. As a retrofit — a bolt-on muzzle programming unit and a fire-control interface — it upgrades existing guns without replacement. The program is phased: AURORA-30 (30 × 165 mm, 2A72-class guns, 150–200 tungsten sub-projectiles per round) proves the timer logic and power-source survival under launch loads before the design is scaled down to AURORA-23 (23 × 152 mm, ZU-23-2-class guns, 50–70 sub-projectiles).

The second line is a kinetic micro-interceptor unmanned aerial vehicle (UAV), another effector node of the kill web (Figure 3): a compact tail-sitting quadrotor of the 7-inch class (177.8 mm airframe height) with four electric propulsion units, vertical launch, and a forward-looking electro-optical camera for the terminal phase. Its effect is purely kinetic, by direct impact, with no explosive warhead — a property that matters in the populated settings of Section 5.4. Vehicles wait at readiness in five-position dock-and-launch modules with charging, health monitoring, and safety interlocks; two modules hold ten ready vehicles. The engagement follows the same accountable sequence as every other effector: ORAO holds the fused track, a designated operator authorizes the launch under the rules of engagement, the mission travels over the authenticated C7ISTAR data link, and the vehicle is guided into the intercept basket and completes the engagement with on-board terminal sensing; on a miss or abort, a commanded parachute-assisted recovery returns it for inspection, recharge, and readiness. A reusable interceptor thus extends the cost-exchange logic of AURORA beyond the gun envelope, engaging the fiber-guided and autonomous platforms that jamming cannot defeat, at a per-engagement cost that scales to mass attack — the reason interceptor drones have become the fastest-growing hard-kill layer internationally, up to the autonomous drone-on-drone interception now reported in the literature [2], [30]. Terminal-guidance performance and the recovery cycle are qualification objectives of the current phase.

The third line is a high-energy laser, past functional-model testing and progressing toward a prototype. Neutralization requires holding a small, agile target in a spot of the order of 10 mm for seconds — a collimator, a rangefinder and a pan-tilt and tracking chain of exceptional precision, all inherited from the VMSIS3 line, whose gyro-stabilization is studied in [31]. All three lines consume what the web already produces — fused track, range and authenticated tasking — and the layered system selects per engagement: soft kill first wherever the threat depends on its radio links, hard kill where soft kill cannot reach.

5.4. Operation in civilian environments

C-UAS operates over and around civilian life, making constraints design inputs rather than afterthoughts:

jamming interacts with regulated spectrum and legitimate users; kinetic and directed-energy effects over populated areas carry collateral risk; and interception of aircraft, however small, engages aviation law and requires a documented decision trail. The operational system therefore provides selective control of jammer power and coverage zones, safeguards against unauthorized triggering unless a fully automatic mode is explicitly enabled, geographic and functional zoning of permitted effects, and evidentiary logging of imagery, tracks and decisions in time-stamped, cryptographically verifiable records [29]. Among hard-kill options, a warhead-free kinetic interceptor with controlled recovery best fits these constraints, introducing neither blast nor fragmentation over the protected area.

6. DEPLOYMENT AND USE CASES

The clearest evidence that the architecture is more than conceptual is that it is deployed. Since 2024 the integrated anti-drone system has protected a major critical-infrastructure facility in an urban environment in Serbia, stopping small commercial and FPV drones before they come within 500 m of the perimeter: sensor towers with panoramic cameras, RF detection and direction-finding sensors and agile VMSIS optics cover roughly one square kilometer, and a countermeasure tower provides RF jamming with selective radiation and coverage control so that the surrounding area is unaffected, all operated through ORAO; because the site is urban, radar was omitted and the passive and optical layers carry detection. The site maps onto concentric functions — early warning at distance, confirmation in the middle zone, countermeasures in the inner zone — under continuous unmanned watch that alerts a small response element.

The same architecture has been engineered for mobility: a configuration on a 4×4 mine-resistant ambush-protected (MRAP) vehicle packages ORAO, the sensor suite, the edge-AI optoelectronic turret, and RF countermeasures onto a protected cross-country platform, rapidly movable to a threatened location and remotely commanded.

The maritime profile extends the architecture to vessels and offshore infrastructure, where operational experience has made the cost-exchange argument unanswerable: in the Red Sea, warships expended roughly 220 air-defense missiles — on the order of a billion dollars — against UAS and missile attacks in a single year [32]. In the configuration engineered for a very large crude carrier (VLCC), two radar heads with overlapping coverage, six acoustic nodes along the hull and four marinized Skua EO/IR turrets feed a shipboard ORAO instance, with two five-position micro-interceptor dock modules on the aft weather deck as the effector layer; the installation is a segregated security enclave with no control path into ship systems and no default external dependency. Maritime sensing experience — surface-wave radar for wide-area sea surveillance [12], [17] and fusion of automatic identification system (AIS) data with over-the-horizon radar tracks [23] — carries directly into this profile, and the low-altitude picture of the vessel federates into the same recognized picture as the shore-side sites.

Temporary sites and public events form a fourth profile: rapidly deployable, soft-kill-preferred and bound by the tightest civilian constraints. Across all profiles the same components scale from a self-contained site to federated regional coverage within the recognized air picture; the passive acoustic node is most valuable in mobile and temporary profiles, where emission control is itself protective.

7. CONCLUSION

There is no single-sensor, single-effector answer to the small-UAS threat: the threat is defined precisely by the properties — small, slow, low, cheap, numerous, increasingly autonomous — that defeat point solutions. The answer is architectural: a distributed multi-modal sensor layer including a passive component effective against non-emitting targets; a classification and fusion stage producing a credible, ranked threat picture at a manageable false-alarm rate; and a kill-web C2 layer connecting any sensor to the best available effector under explicit rules of engagement — the shortest defensible path from sensor to shooter, under human authority and recorded at every step.

The presented results substantiate each element: ARES discriminates drones from birds, aircraft, and helicopters at the sensor edge and fuses tracks into one picture; the agile optoelectronic turret confirms a cued target in under two seconds; the passive acoustic sensor addresses the radio-silent platforms that defeat the RF layer; and ORAO binds these sensors with the operational jammer and with the airburst gun, micro-interceptor, and laser under development into one kill web, whose operational core has protected national critical infrastructure since 2024. Future work covers integrating the acoustic layer into the wider air picture, maturing AURORA, qualifying the micro-interceptor and its recovery cycle, advancing the high-energy laser, and improving the adversarial robustness of classification; defense against coordinated multi-axis swarms remains the next open problem [3], [5]. Sensors, AI and effectors are becoming one sovereign loop, in which the decisive quantity is the time — and the trust — with which an observation becomes an effect.

Acknowledgment

The Vlatacom Institute of High Technologies funded this work through its projects VMSIS3_Advanced and AI-DECIDE. Passive acoustic sensing was developed jointly with the School of Electrical Engineering, University of Belgrade.

References

[1] HOLLENBECK,N., ALTAF,M.H., AVILA,F., RAMIREZ,J., SHARMA,A., JENSEN,B.: *Calculating the Cost-Effectiveness of Russia's Drone Strikes*, Center for Strategic and International Studies, Washington, DC, 2025.

[2] WATLING,J., BRONK,J.: *Protecting the Force from Uncrewed Aerial Systems*, RUSI Occasional Paper, Royal United Services Institute, London, 2024.

[3] PARK,S., KIM,H.T., LEE,S., JOO,H., KIM,H.: *Survey on Anti-Drone Systems: Components, Designs, and Challenges*,

IEEE Access, 9 (2021) 42635–42659.

[4] GUVENC,I., KOOHIFAR,F., SINGH,S., SICHITIU,M.L., MATOLAK,D.: *Detection, Tracking, and Interdiction for Amateur Drones*, IEEE Communications Magazine, 56(4) (2018) 75–81.

[5] XU,L., LUO,Z.: *Anti-UAV Detection and Identification Technology: Fundamentals, Methods and Challenges*, Physical Communication, 71 (2025) 102676.

[6] U.S. DEPARTMENT OF DEFENSE: *Strategy for Countering Unmanned Systems*, Washington, DC, 2024.

[7] VULIC,I.: *Introduction to Command-Information Systems*, in: Command-Information Systems, University of Defence, Belgrade, 2026 (in Serbian).

[8] VUJIC,S., PERIC,D., LIVADA,B., RADISAVLJEVIC,M., DOMAZET,D.: *Optical Protective Window Design and Material Selection Issues in the Multi-Sensor Electro-Optical Surveillance Systems*, Sensors, 23 (2023) 7211.

[9] PERIC,D., LIVADA,B., PERIC,M., VUJIC,S.: *Thermal Imager Range: Predictions, Expectations, and Reality*, Sensors, 19 (2019) 3313.

[10] PAVLOVIC,M., BANJAC,Z., KOVACEVIC,B.: *Object Tracking in SWIR Imaging Based on Both Correlation and Robust Kalman Filters*, IEEE Access, 11 (2023).

[11] CIZELJ,D., PAVLOVIC,M., UNKASEVIC,T., BANJAC,Z.: *Automated Target Tracking Activation Based on Motion Detection*, Proc. 10th Int. Conf. on Electrical, Electronic and Computing Engineering (IcETRAN), East Sarajevo, 2023.

[12] NIKOLIC,D., STOJKOVIC,N., PUZOVIC,S., POPOVIC,Z., STOJILJKOVIC,N., GRBIC,N., ORLIC,V.: *Increasing Maritime Safety and Security in the Off-Shore Activities with HFSWRs as Primary Sensors for Risk Assessment*, Journal of Marine Science and Engineering, 11 (2023).

[13] PETROVIC,P., GRBIC,N., LEKIC,N., PERIC,M., SAVIC,S.: *Aircraft Detection Range for HF Radar*, Proc. 31st Telecommunications Forum (TELFOR), Belgrade, 2023.

[14] POPADIC,I., PERIC,M., TOMIC,B., SIMIC,A., MILANOVIC,P.: *High-End Video Processing Platform for Electro-Optical Systems*, Proc. 9th International Scientific Conference on Defensive Technologies (OTEH), Belgrade, 2020.

[15] FARLIK,J., KRATKY,M., CASAR,J., STARY,V.: *Multispectral Detection of Commercial Unmanned Aerial Vehicles*, Sensors, 19(7) (2019) 1517.

[16] TAHA,B., SHOUFAN,A.: *Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research*, IEEE Access, 7 (2019) 138669–138682.

[17] GOLUBOVIC,D., ERIC,M., VUKMIROVIC,N., ORLIC,V.: *High-Resolution Sea Surface Target Detection Using Bi-Frequency High-Frequency Surface Wave Radar*, Remote Sensing, 16 (2024).

[18] KASCA,L., PAVLOVIC,D.S., BJELIC,M., NESKOVIC,DJ., MARKOVIC,M.: *Underwater Model for Acoustic Waves Propagation*, Proc. 23rd International Symposium INFOTEH-JAHORINA, IEEE, 2024.

[19] GHENESCU,M.T., GHENESCU,V., CARATA,S.V.: *Acoustic Source Drone Detection System Using Tetrahedral Microphone Array and Deep Neural Networks*, Sensors, 26(6) (2026) 1778.

[20] PRODANOVIC,R., RANCIC,D., VULIC,I., ZORIC,N., BOGICEVIC,D., OSTOJIC,G., SARANG,S., STANKOVSKI,S.: *Wireless Sensor Network in Agriculture: Model of Cyber Security*, Sensors, 20(23) (2020) 6747.

[21] PRODANOVIC,R., SARANG,S., RANCIC,D., VULIC,I., STOJANOVIC,G.M., STANKOVSKI,S., OSTOJIC,G., BARANOVSKI,I., MAKSOVIC,D.: *Trustworthy Wireless Sensor Networks for Monitoring Humidity and Moisture Environments*, Sensors, 21(11) (2021) 3636.

[22] HALL,D.L., LLINAS,J.: *An Introduction to Multisensor Data Fusion*, Proceedings of the IEEE, 85(1) (1997) 6–23.

[23] MOSTAFA,M.A.R.M., VUCETIC,M., STOJKOVIC,N., LEKIC,N., MAKAROV,A.: *Fuzzy Functional Dependencies as a Method of Choice for Fusion of AIS and OTHR Data*, in: Smart Sensors and Devices in Artificial Intelligence, Springer, Basel, 2022.

[24] ILIC,N., VUCETIC,M., MAKAROV,A., PETROVIC,R., PUNT,M.: *Adaptive Asynchronous Gossip Algorithms for Consensus in Heterogeneous Sensor Networks*, IEEE Internet of Things Journal, 12 (2025).

[25] CLARK,B., PATT,D., SCHRAMM,H.: *Mosaic Warfare: Exploiting Artificial Intelligence and Autonomous Systems to Implement Decision-Centric Operations*, Center for Strategic and Budgetary Assessments, Washington, DC, 2020.

[26] CLARK,B., PATT,D., SERAPHIN,A., MILES,W.: *Integrated by Mission — Federated for Execution*, Hudson Institute, Washington, DC, 2024.

[27] VULIC,I., PRODANOVIC,R., TOT,I., BOGICEVIC,D.: *Model for Authenticating the Internet of Military Things and Internet of Battlefield*, Proc. 10th Int. Conf. on Information Society and Technology (ICIST), Kopaonik, 2020, 206–210.

[28] PRODANOVIC,R.I., VULIC,I.B.: *Model for PKI Interoperability in Serbia*, Vojnotehnicki glasnik / Military Technical Courier, 65(2) (2017) 530–549.

[29] VULIC,I., PRODANOVIC,R., VUKCEVIC,G., SRETENOVIC,S.: *Trust Establishing Model in IoT Using PKI and Timestamp*, Proc. 8th Int. Conf. on Information Society and Technology (ICIST), Kopaonik, 2018, 333–338.

[30] ROTHE,J., KESSLER,N., HENRIQUEZ WEHR,M., HOHBACH,A., STROHMEIER,M., MONTENEGRO,S.: *Autonomous Drone-on-Drone Interception Using an Integrated LiDAR–Vision Detection System for High-Precision Capture*, Drones, 10(6) (2026) 420.

[31] MILANOVIC,P., POPADIC,I., KOVACEVIC,B.: *Gyroscope-Based Video Stabilization for Electro-Optical Long-Range Surveillance Systems*, Sensors, 21 (2021) 6219.

[32] PYLE,F.: *Operations in the Red Sea: Lessons for Surface Warfare*, Center for Strategic and International Studies, Washington, DC, 2024.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AUTONOMOUS LOCOMOTION OF ROBOTIC DOGS IN MILITARY APPLICATIONS

NINA MITRIČEVIĆ
Military Technical Institute, Ministry of Defence, Ratka Resanovića 1, Belgrade, nmitricevic@gmail.com

RADE PAVLOVIĆ
Military Technical Institute, Ministry of Defence, Ratka Resanovića 1, Belgrade, rade.pavlovic@mod.gov.rs

Abstract: *The rapid development of robotics, artificial intelligence, computer vision, and autonomous navigation has significantly expanded the potential applications of unmanned ground systems in military environments. Among the most interesting developments are quadrupedal robots, commonly known as robotic dogs. Unlike traditional wheeled or tracked unmanned ground vehicles, quadrupedal robots are designed to move across complex terrain while maintaining stability and adapting their walk to environmental conditions. Their ability to operate without continuous direct human control makes them particularly relevant for reconnaissance, surveillance, logistics, hazardous-area inspection, communications support, and other military tasks. Autonomous locomotion is one of the most important technological capabilities of a robotic dog. It involves the integration of sensors, perception systems, localization algorithms, path planning, obstacle avoidance, motion control, and artificial intelligence. The robot must be capable of understanding its surroundings, determining its position, selecting a safe route, and controlling its legs in real time. These requirements become considerably more difficult in military environments, where terrain can be unpredictable, communication links can be unreliable, and environmental conditions may change rapidly. This paper examines the principles behind autonomous locomotion in robotic dogs and discusses their potential military applications. Particular attention is given to navigation technologies, sensor fusion, artificial intelligence, locomotion control, operational advantages, technical limitations, cybersecurity, human supervision, and ethical considerations. The paper also considers the future development of quadrupedal military robots and the importance of maintaining meaningful human control over autonomous systems.*

Keywords: *robotic dogs, autonomous locomotion, military robotics, unmanned ground vehicles, autonomous systems.*

1. INTRODUCTION

Robotics has become an increasingly important area of technological development in the defense sector. Modern military forces operate in environments that can be dangerous, unpredictable, and difficult for human personnel to access [1]. Unmanned systems provide an opportunity to perform certain tasks while reducing the exposure of soldiers to physical hazards. Although unmanned aerial vehicles have received considerable attention, unmanned ground robots are also becoming increasingly capable.

One particularly interesting category of unmanned ground systems is the quadrupedal robot, often referred to as a robotic dog. These machines use four legs instead of wheels or continuous tracks. This configuration allows them to negotiate terrain that may be difficult for conventional ground vehicles, including stairs, uneven surfaces, rubble, narrow passages, and natural terrain [2].

The military interest in robotic dogs is not based only on their ability to walk. Their greater significance comes from their potential to combine mobility with advanced perception and autonomy [3]. A robot that can independently navigate through an environment can perform tasks without requiring an operator to control every individual movement. Instead, the operator can provide a higher-level objective while the robot handles many aspects of movement and navigation autonomously.

Autonomous locomotion is therefore a multidisciplinary problem. It combines mechanical engineering, robotics, artificial intelligence, computer vision, control theory, embedded computing and sensor technology. A successful system must continuously perceive its surroundings, estimate its own position, construct an understanding of the terrain, determine an appropriate path, and coordinate its legs to maintain balance [4].

In military applications, these challenges are particularly significant. A robotic dog may need to operate in darkness, dust, rain, snow, confined spaces or environments where satellite navigation is unavailable. They can also temporarily lose communication with its human operator. Consequently, autonomous operation must be designed to remain predictable and safe under uncertain conditions.

The objective of this paper is to examine the technological foundations and potential military applications of autonomous locomotion in robotic dogs. The discussion focuses on mobility, navigation, sensing, artificial intelligence, operational advantages, limitations, and others.

Corresponding author: R. Pavlović;
E-mail: rade.pavlovic@mod.gov.rs
DOI: 10.5937/OTEH26075M

2. AUTONOMOUS LOCOMOTION

The idea of using four-legged machines for transportation and military purposes (figure 1) is not new. Early robotic systems were generally limited by mechanical complexity, inadequate computing power and the lack of sufficiently advanced sensors [3]. A four-legged robot must coordinate several joints simultaneously while maintaining balance, which makes its control problem significantly more complicated than that of a simple wheeled platform [5].

Figure 1. Four-legged robot.

Advances in electric motors, batteries, embedded processors, inertial sensors, cameras, and machine-learning algorithms have changed this situation. Modern quadrupedal robots can perform dynamic movements that were previously extremely difficult for machines. They can walk, turn, climb stairs, recover from disturbances, and adapt their movements to different surfaces [6].

A major advantage of the quadrupedal configuration is the ability to maintain contact with the ground at several points. Depending of the moving, the robot can distribute its weight across multiple legs and adjust its center of mass. This is useful when operating on uneven terrain.

The mechanical design of a robotic dog generally includes four legs with multiple degrees of freedom. Each leg may contain several actuated joints, allowing the foot to be positioned in three-dimensional space. Sensors within the joints provide information about position and movement, while inertial sensors provide information about the orientation and acceleration of the robot.

2.1. Sensors and sensor fusion

One of the major advantages of quadrupedal robots is their ability to negotiate irregular terrain. However, this ability requires the robot to understand what type of terrain lies in front of it. A robotic dog may encounter stairs, rocks, holes, rubble, mud, vegetation, cables, walls, doors, and other obstacles. Each situation requires a different movement strategy. For example, a flat floor allows the robot to use a regular walking pattern. A staircase requires the robot to raise its feet to a different height and coordinate its legs appropriately. Rubble may require slower movement and continuous adjustment of foot placement.

Terrain perception can be based on depth information obtained from cameras or LiDAR. The robot can construct a three-dimensional representation of the environment and identify areas that appear suitable or unsuitable for walking [7].

Modern platforms may also incorporate cameras, sensors, depth cameras, microphones, thermal sensors, and other instruments (figure 2). These systems transform the robotic dog from a simple walking machine into a mobile autonomous sensing platform.

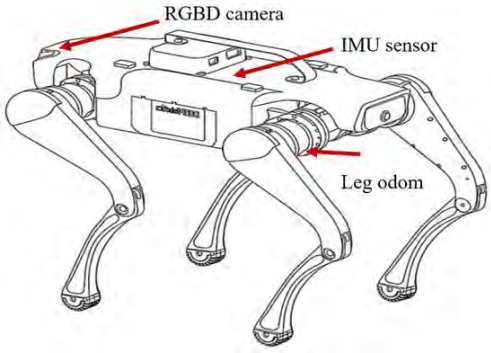


Figure 2. Sensors on four-legged robot.

Perception is the process through which the robot collects and interprets information about the environment. Cameras provide visual information, while LiDAR systems can provide three-dimensional measurements of surrounding objects and terrain. Inertial Measurement Units (IMUs) measure acceleration and angular velocity. These measurements help determine whether the robot is tilting, accelerating, or rotating. Joint encoders provide information about the positions of the robot's legs. No individual sensor is perfect, cameras can be affected by darkness, smoke, dust, or poor visibility, LiDAR can be affected by certain environmental conditions and may provide incomplete information and Inertial sensors accumulate measurement errors over time.

For this reason, modern autonomous robots frequently use sensor fusion [8]. Sensor fusion combines information from several sensors to produce a more reliable representation of the robot and its surroundings.

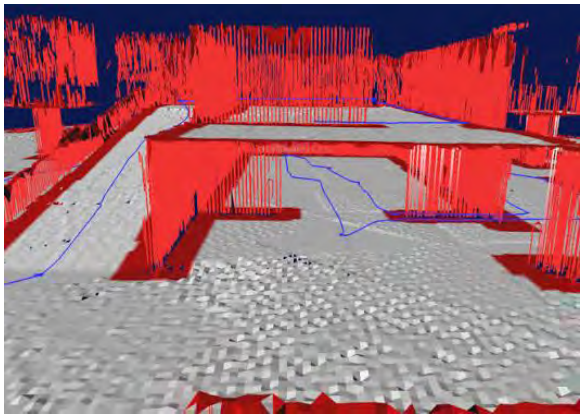


Figure 3. Traversability map example.

An important concept is the traversability map (figure 3). Instead of treating the environment simply as a collection of obstacles, the robot can assign different levels of

difficulty or risk to different areas. A flat surface may be considered highly traversable, while a steep or unstable surface may be considered less desirable.

2.2. Artificial intelligence in autonomous locomotion

Autonomous locomotion refers to the ability of a robot to move through an environment with limited or no continuous human control. In the case of a robotic dog, autonomous locomotion involves several interconnected processes. First, the robot must perceive its environment. It needs to identify obstacles, terrain characteristics, and possible paths. Second, it must estimate its own position and orientation. Third, it must determine a route to its destination. Finally, it must transform the selected route into physical movements of its legs. The robot can then use this information during path planning.

Obstacle avoidance is another essential component. When the robot detects an obstacle, it can modify its planned trajectory. In a simple environment, this may involve moving around the obstacle. In a more complex environment, the system may need to evaluate several possible routes. The ability to perform obstacle avoidance autonomously reduces the workload of the human operator. Instead of manually controlling every movement, the operator can monitor the robot and intervene when necessary.

Artificial intelligence plays an increasingly important role in autonomous robotics [9]. Traditional robotic control systems rely heavily on mathematical models and predefined rules. These approaches remain important, but machine learning can provide additional capabilities. Neural networks can be trained to recognize objects, classify terrain, and interpret visual information. A robot can potentially distinguish between different types of surfaces and identify objects that may affect its movement.

Machine learning can also be applied to locomotion. A robot can learn how to coordinate its joints to achieve stable movement under different conditions. Reinforcement learning, for example, allows an agent to improve its behaviour through repeated interaction with a simulated or physical environment.

Simulation is particularly valuable for this purpose. Training a physical robot through millions of failed attempts would be expensive and potentially damaging. Instead, researchers can create virtual environments in which a robot learns different movement strategies. Successful strategies can then be transferred to a physical system, although differences between simulation and reality must be carefully addressed.

Artificial intelligence can also contribute to adaptive control. If the robot detects that a surface is slippery or uneven, its control system can modify its movement accordingly. Nevertheless, artificial intelligence should not be treated as a perfect solution. Machine-learning systems can make mistakes, particularly when they encounter situations that differ significantly from their training data. Military systems therefore require careful validation, testing, monitoring, and human oversight.

3. AUTONOMOUS NAVIGATION IN MILITARY ENVIRONMENTS

Military environments present several challenges that are less common in civilian robotic applications. Terrain may be damaged, visibility may be poor, communications may be unreliable, and the robot may need to operate for extended periods without direct supervision.

Urban environments are especially complicated like buildings rooms, corridors, staircases, doorways, narrow passages and numerous obstacles (figure 4). A robot must be able to construct and update a map while moving through such spaces. Natural environments create different challenges, forests, fields, mountains, and areas affected by natural disasters may contain irregular ground surfaces and unpredictable obstacles.

Figure 4. Robot dog moving down the stairs.

A major advantage of quadrupedal robots is their ability to adapt to terrain without requiring prepared roads. This makes them potentially useful for reconnaissance and inspection in areas where conventional vehicles cannot easily operate. Another important characteristic is the possibility of autonomous return behaviour (back to HOME). If a robot loses communication with its operator, an appropriately designed system could enter a predefined safe state or attempt to return to a known location. Such functions are important because autonomous systems should not become unpredictable when communication is interrupted.

However, autonomous navigation does not necessarily mean completely independent operation. A more realistic approach for many military applications is **human-supervised autonomy**. In this model, the robot performs low-level navigation autonomously while a human remains responsible for higher-level decisions.

3.1. Military Applications

Robotic dogs can potentially support a wide range of military activities. Many of these applications do not require the robot to carry weapons. Instead, their value comes from mobility, sensing, and the ability to operate in hazardous areas. One of the most obvious applications is reconnaissance [10]. A robotic dog equipped with cameras and other sensors can investigate an area while allowing personnel to remain at a safer distance (figure 5).

The robot may transmit visual information to an operator, helping personnel understand the physical characteristics of an unfamiliar environment. Autonomous navigation can allow the robot to move through an area while the operator focuses on interpreting sensor information. This can reduce the amount of direct manual control required.

Figure 5. Robotic dog in reconnaissance application.

Robotic dogs can also be useful for inspecting potentially dangerous locations (figure 6), like damaged buildings, industrial facilities, areas affected by explosions, and environments containing unknown hazards.

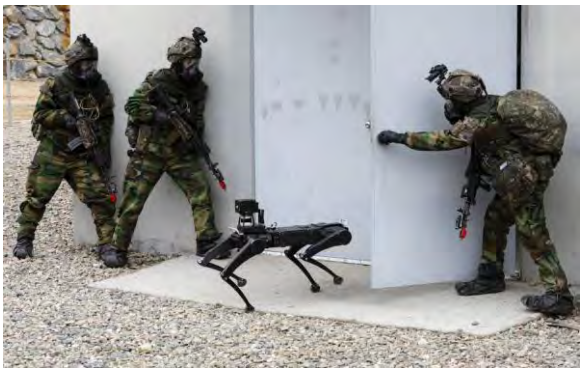


Figure 6. Using of Robot dogs in potentially dangerous locations.

A robot can enter an area that may be unsafe for humans and collect information using cameras or other sensors. The primary benefit is risk reduction. Instead of immediately exposing a person to an unknown environment, personnel can first use an unmanned platform to gather information.

Another possible application is transportation of equipment (figure 7). Quadrupedal robots can potentially carry supplies across terrain that is difficult for wheeled systems.

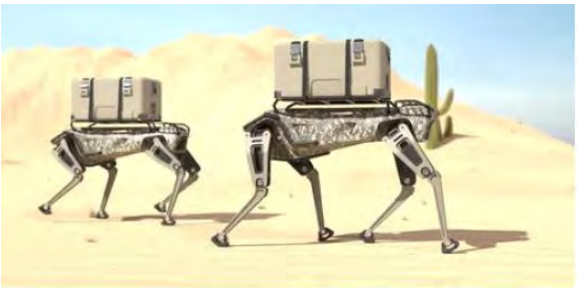


Figure 7. Robotic dogs in logistic application.

Their ability to follow a person or navigate toward predefined locations could make them useful as robotic pack animals. The practical usefulness of this application depends heavily on battery capacity, payload, reliability, and terrain. A robot that can carry equipment for only a short period may have limited value compared with conventional transportation methods.

A robotic dog equipped with communication equipment could potentially serve as a mobile communications node. Its mobility could allow it to reposition itself when required. Such systems could be useful in environments where fixed communication infrastructure is damaged or unavailable. However, communications equipment adds weight and consumes energy, so system designers must balance communication capabilities against mobility and endurance.

Although search and rescue is not exclusively a military mission, armed forces frequently participate in disaster response and emergency operations. A quadrupedal robot can potentially search damaged structures or difficult terrain while carrying cameras and sensors. Autonomous navigation can help it cover areas systematically.

In this context, the robot's ability to climb obstacles and maintain balance may provide advantages over conventional wheeled platforms.

4. CONCLUSION

Autonomous locomotion represents one of the most important capabilities of modern quadrupedal robots. By combining advanced mechanical systems with sensors, artificial intelligence, localization, mapping, path planning, and control algorithms, robotic dogs can move through environments that are difficult for traditional ground vehicles.

Military applications include reconnaissance, surveillance, hazardous-area inspection, logistics, communications support, and search and rescue. In many of these roles, the most significant benefit is the ability to perform useful tasks while reducing the exposure of personnel to dangerous environments.

At the same time, robotic dogs remain subject to important technical limitations. Battery endurance, environmental conditions, communication reliability, localization errors, maintenance requirements, cost, and cybersecurity all affect their practical usefulness.

Autonomy also raises ethical and legal questions, particularly when autonomous mobility is combined with other military capabilities. Maintaining meaningful human control over consequential decisions is therefore an important principle for responsible development.

The future of robotic dogs will probably not involve simply making machines more autonomous. Successful systems will combine autonomy with reliability, human supervision, cybersecurity, explainability, and robust performance in unpredictable environments.

Autonomous quadrupedal robots represent an important step in the development of unmanned ground systems. Their ability to combine mobility, sensing, and intelligent

navigation makes them potentially valuable tools for military and emergency applications. However, technological capability should be developed alongside appropriate human oversight and responsible operational policies. The ultimate value of robotic dogs will depend not only on how well they can walk, but also on how safely, reliably, and responsibly they can be integrated into human-centered systems.

The future development of robotic dogs will likely involve improvements in several areas. The first one is energy efficiency. More efficient motors and batteries could significantly increase operational endurance. The second is artificial intelligence. Improvements in perception and machine learning could allow robots to better understand complex environments. The third is sensor technology. Smaller, lighter, and more capable sensors will allow robots to collect more information without a corresponding increase in size or weight.

Another important development is multi-robot cooperation. Several robots could potentially share information about an environment and coordinate their movement. Such systems could improve coverage and reduce the dependence on a single platform. However, increasing the number of autonomous systems also increases the complexity of coordination and cybersecurity.

Future robotic dogs may also become more modular. Different equipment packages could allow the same basic platform to perform different roles, such as reconnaissance, logistics, communications support, or search and rescue.

The most important development, however, may be improved reliability. Military users require systems that behave predictably under difficult conditions. A robot that performs impressively in a laboratory but fails in dust, rain, or damaged terrain has limited operational value.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] SIMIĆ, N., STANKOVIĆ, A., MARKOVIĆ, D.: The role of drones and robotic dogs in tactical supply chains using a model for logistical support planning. The tenth international conference for transport and logistic, *environments*, *11*, 8 (2025).

[2] SU, X., SHI, G., ZHONG, J., LI, Y., DAI, W., XU, A. G., YAN, Z.: The implementation of robotic dogs in automatic detection and surveillance of red imported fire ant nests. *Pest Management Science*, *80*(10), 5277-5285, (2024).

[3] MADHUSHREE, K., KEERTHANA, M., POORNIMA, J., KAVYA, R., PADMAVATHI, M., PAVITHRA, G., MANJUNATH, T. C.: Development of a highly capable robotic dog with enhanced sensory and communication abilities for military applications. In *AIP Conference Proceedings* (Vol. 3149, No. 1, p. 140029). AIP Publishing LLC, (2024).

[4] ZHAI, C., ZHANG, P., XU, H., YUAN, X., ZHOU, L., WU, R.: The application and inspiration of robots in the US military. In *8th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)* (pp. 24-28), IEEE, (2023).

[5] RAIBERT, M., BLANKESPOOR, K., NELSON, G., PLAYTER, R.: Big Dog, the Rough-Terrain Quadruped Robot. *Proceedings of the 17th World Congress, International Federation of Automatic Control*, (2008).

[6] KALAKRISHNAN, M., BUCHLI, J., PASTOR, P., MISTRY, M., SCHAAL, S.: Learning, Planning, and Control for Quadruped Locomotion over Challenging Terrain. *The International Journal of Robotics Research*, 30(2), 236–258, (2011).

[7] ZOU, H., ZHOU, C., LI, H., WANG, X., WANG, Y.: Application of 3D point cloud and visual-inertial data fusion in Robot dog autonomous navigation. *Plos one*, *20*(2), e0317371, (2025).

[8] CHEN, T., HUANGFU, Y., SRIGRAROM, S., KHOO, B. C.: Path planning and motion control of robot dog through rough terrain based on vision navigation. *Sensors*, *24*(22), 7306, (2024).

[9] WAGNER, W., VIIDALEPP, A., IDOIAGA-MONDRAGON, N., TALVES, K., LILLEMAEE, E., PEKAREV, J., OTSUS, M.: Lay representations of artificial intelligence and autonomous military machines. *Public Understanding of Science*, *32*(7), 926-943, (2023).

[10] MUDA, N. R. S., FANDI, P., & AZIZIA, P.T.: Design and Build a Dog Paw-shaped Robot Camouflage for Reconnaissance Equipped with Ss2 V2 Weapon. *International of IJNRSM*, *1*(1), 641-653, (2025).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

UAV POSITION ESTIMATION THROUGH IMAGE MATCHING BETWEEN DRONE IMAGERY AND MAPS

RADE PAVLOVIĆ
Military Technical Institute, Ministry of Defence, Ratka Resanovića 1, Belgrade, rade.pavlovic@mod.gov.rs

SARA PAVLOVIĆ
Union University – School of Computing, 6/VI Kneza Mihaila, Belgrade, sarapavlovic1977@gmail.com

Abstract: *Accurate position estimation is a fundamental requirement for the autonomous navigation of unmanned aerial vehicles (UAVs). Most modern drones rely on Global Navigation Satellite Systems (GNSS), such as GPS, to determine their geographical position. However, satellite navigation can become unreliable in environments with signal obstruction, interference, or poor satellite visibility. An alternative approach is visual localization, which estimates the position of a drone by comparing images captured by its onboard camera with reference maps or georeferenced satellite imagery. This paper presents the principles and methods of UAV position estimation through image matching between drone imagery and maps. The main stages of the process include image acquisition, preprocessing, feature extraction, image matching, geometric transformation, and geographical coordinate estimation. Particular attention is given to traditional computer vision techniques, such as Scale-Invariant Feature Transform (SIFT), Oriented FAST and Rotated BRIEF (ORB), and feature matching using geometric models. The paper also discusses deep learning-based approaches, image registration, and the challenges associated with differences in viewing angle, image resolution, illumination, and environmental conditions. Finally, evaluation metrics and possible applications are examined. Visual localization can provide an additional source of positioning information and may complement conventional navigation systems. Its effectiveness, however, depends on the quality of reference maps, the reliability of image correspondences, and the accuracy of the geometric transformation.*

Keywords: *unmanned aerial vehicle, visual localization, image matching, satellite imagery, geographical positioning, computer vision, image registration, GNSS-denied navigation.*

1. INTRODUCTION

Unmanned aerial vehicles have become increasingly important in modern engineering, scientific research, and commercial applications. They are used for aerial photography, agricultural monitoring, infrastructure inspection, environmental observation, surveying, and search-and-rescue operations. Many of these applications require accurate knowledge of the UAV's geographical position.

The Global Navigation Satellite System is the most common source of positioning information for modern drones. By receiving signals from navigation satellites, a UAV can estimate its latitude, longitude, and altitude. This information is used by the flight controller to maintain a planned trajectory and execute autonomous missions.

Despite its widespread use, satellite-based positioning has several limitations. Buildings, mountains, and other obstacles can obstruct satellite signals. Signal reflections may introduce positioning errors, while interference can reduce the availability or reliability of navigation information. These limitations have encouraged research into alternative localization techniques, including inertial navigation, visual odometry, and image-based geolocation [1].

One of approaches is to determine the UAV's position by comparing an image captured by its onboard camera with a reference map. The reference image may be obtained from satellite imagery, aerial photographs, or a previously constructed georeferenced map. If a sufficiently accurate correspondence is established between the two images, the geographical location of the camera can be estimated.

This process is commonly referred to as visual localization or visual geolocalization. In aerial applications, it often involves matching images captured from a relatively low altitude with satellite images acquired from a much higher viewpoint.

The main objective of this paper is to explain the theoretical foundations and practical implementation of image-based UAV position estimation. The paper describes the image-matching process, presents commonly used computer vision algorithms, discusses geographical coordinate estimation, and examines the main factors influencing localization accuracy.

2. THEORETICAL FOUNDATIONS OF VISUAL LOCALIZATION

Visual localization is the process of determining the position of a camera or mobile platform by analyzing visual information. In the case of a UAV, the system

Corresponding author: R, Pavlović;
E-mail: rade.pavlovic@mod.gov.rs
DOI: 10.5937/OTEH26076

attempts to identify the geographical location corresponding to the area visible in the camera image [2].

The general principle is based on comparing two sources of visual information (figure 1):

- query image captured by the UAV's onboard camera.
- reference image with known geographical coordinates.

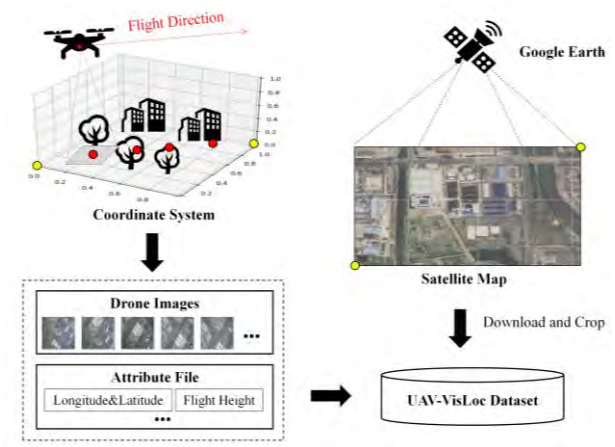


Figure 1. The dataset collecting process [3].

The query image contains visual features of the terrain, buildings, roads, fields, rivers, and other objects. The reference map contains corresponding geographical structures. By identifying common features, the system can estimate where the UAV is located.

The reference map must be georeferenced, meaning that its image coordinates are associated with real-world geographical coordinates. These coordinates may be expressed as latitude and longitude or as projected coordinates, such as easting and northing.

The localization process can be divided into two principal tasks. The first is identifying the correct geographical area represented in the UAV image. The second is estimating the precise alignment between the UAV image and the reference map [4].

The first task is often described as image retrieval or place recognition, while the second involves image matching and geometric registration. Combining these stages creates a coarse-to-fine localization pipeline.

2.1. Coordinate Systems

Accurate position estimation requires an understanding of the coordinate systems involved. A digital image uses a two-dimensional pixel coordinate system. A pixel is typically represented by coordinates ((u,v)), where (u) represents the horizontal position and (v) represents the vertical position in the image.

A map, in contrast, represents locations in a geographical coordinate system. In the WGS 84 system, a location is commonly expressed using latitude and longitude. Projected coordinate systems represent locations using linear coordinates, such as metres [5].

The purpose of image registration is to establish a mathematical relationship between these coordinate systems. For example, if a pixel in the UAV image can be matched to a known point in a georeferenced map, its corresponding geographical location can be determined. Multiple reliable correspondences can be used to estimate the transformation between the images.

However, the camera's position is not always identical to the geographical coordinates of the matched ground point. The camera is located above the Earth's surface, and its viewing direction affects the relationship between image coordinates and ground coordinates. Camera altitude, orientation, and terrain elevation must therefore be considered when estimating the UAV's actual position.

2.2. Absolute and Relative Localization

Two important localization concepts are absolute localization and relative localization. Absolute localization estimates the UAV's position within a global or geographically referenced coordinate system. Matching a camera image to a georeferenced satellite map is an example of absolute visual localization [6].

Relative localization estimates the movement of the UAV with respect to a previous position or observation. Visual odometry, for example, estimates motion by analyzing changes between consecutive camera frames.

Relative localization can provide useful motion information, but errors may accumulate over time. Absolute localization can correct this drift by connecting the current observation to a reference map.

A system that combines both approaches can use visual odometry to estimate movement between frames and map matching to periodically correct its geographical position.

3. IMAGE MATCHING ALGORITHMS

Feature-based image matching is one of the traditional approaches to visual localization. Instead of comparing every pixel directly, the algorithm identifies distinctive points in each image and establishes correspondences between them. A feature point may represent a corner, intersection, or other locally distinctive image structure. Each point is associated with a descriptor that describes its surrounding visual pattern.

The matching algorithm compares descriptors from the UAV image with descriptors from the reference map (figure 2). Similar descriptors indicate potential correspondences. Feature-based approaches are particularly useful when the images differ in scale, rotation, or moderate changes in illumination.

Figure 2. Drone images and satellite map from dataset [3].

Several algorithms have been developed for this purpose, including SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), and ORB (Oriented FAST and Rotated BRIEF). Their performance depends on image characteristics, computational resources, and the degree of difference between the two viewpoints.

3.1. Scale-Invariant Feature Transform (SIFT)

The Scale-Invariant Feature Transform is a widely used algorithm for detecting and describing local image features [7]. SIFT identifies distinctive keypoints at different image scales (figure 3). It then constructs descriptors based on the local gradient orientations surrounding each keypoint.

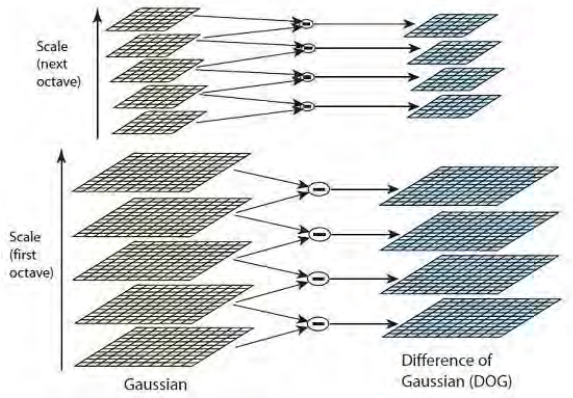


Figure 3. Diagram of SIFT process [8].

One of the principal advantages of SIFT is its relative robustness to changes in image scale and rotation. This makes it useful when a UAV image and a reference map represent the same geographical area at different resolutions. For example, a road intersection may appear relatively large in a low-altitude drone image and much smaller in a satellite image. SIFT can identify corresponding local structures even when their apparent sizes differ.

However, SIFT can require considerable computational resources when processing high-resolution images or large collections of map tiles. It may also produce unreliable matches in areas with repetitive patterns, such as similar rooftops or agricultural fields.

3.2. ORB Feature Detection

ORB combines the FAST keypoint detector with the BRIEF descriptor and an orientation estimation procedure [9]. Compared with traditional floating-point descriptors, ORB uses binary descriptors that can be matched efficiently using Hamming distance. This makes ORB attractive for real-time applications and embedded computing platforms. It can process images relatively quickly and is available in widely used computer vision libraries.

However, ORB may be less robust than some alternatives when there are substantial changes in scale, viewpoint, or image appearance (figure 4). Its effectiveness therefore depends on the particular environment and camera configuration.

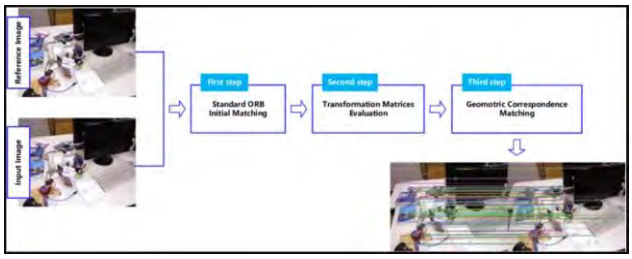


Figure 4. The framework of Oriented FAST and Rotated BRIEF [10].

For a UAV operating on a limited computational budget, ORB can provide a useful starting point. Its performance should nevertheless be evaluated against the expected range of flight altitudes and viewing angles.

3.3. Deep learning Matching

Deep learning provides an alternative approach by learning image representations directly from large datasets. Instead of relying exclusively on manually designed descriptors, a neural network can learn features that are more robust to changes between the drone image and the reference map [11]. A deep learning matcher can therefore be designed to extract high-level semantic and geometric information from both images and estimate their similarity or spatial correspondence (figure 5).

The general architecture of a deep learning-based image-to-map matching system consists of three main stages: feature extraction, feature matching, and geographical localization [12]. The first stage generates compact feature representations from the drone image and the map. The second stage establishes correspondences between the extracted representations. Finally, the matched information is transformed into an estimated geographical position.

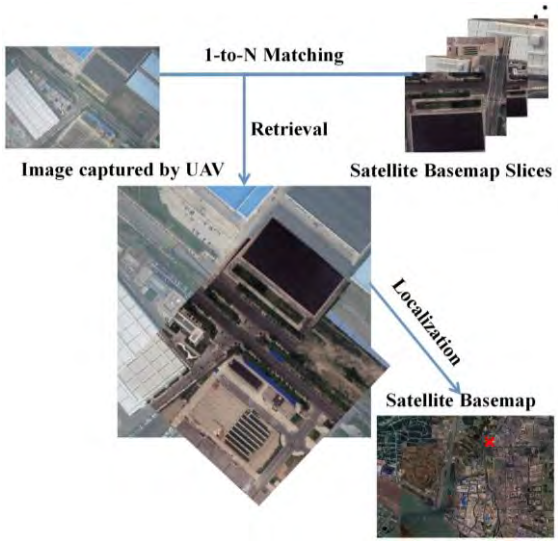


Figure 5. General architecture of a deep learning-based image-to-map matching system.

A major advantage of deep learning-based matching is its ability to learn robust representations directly from data. The learned features can be more tolerant to variations in illumination, viewpoint, scale, and scene appearance than traditional hand-crafted descriptors.

Another advantage is the possibility of combining global and local information. Global descriptors allow efficient candidate retrieval, while local matching provides accurate spatial correspondence.

However, several limitations remain. Training a reliable model requires a sufficiently large and representative dataset. Performance can decrease when the deployment environment differs significantly from the training data. Large changes in terrain, construction, vegetation, weather, or seasonal appearance may reduce the visual similarity between the drone image and the reference map.

Computational requirements are another important consideration. Deep neural networks can require significant GPU resources during both training and inference. For real-time UAV applications, model compression, quantization, efficient architectures, and hardware acceleration may therefore be necessary.

4. EXPERIMENTAL RESULTS AND POSITIONING ERROR ANALYSIS

To evaluate the effectiveness of image-based UAV localization, experiment is considered in which drone images are matched against a georeferenced satellite map. The objective is to estimate the geographical position of the UAV and measure the difference between the estimated coordinates and the known reference position. In this study dataset containing 200 UAV images is assumed. The images represent different camera orientations, flight altitudes, and environmental conditions. Each image is associated with a reference position obtained from an independent positioning system.

Three image-matching approaches are considered: SIFT, ORB, and deep learning-based feature matcher. All three methods are evaluated using the same image set and reference map. The experiment assumes that the reference map is accurately georeferenced and that the ground-truth positions are expressed in a suitable projected coordinate system, with distances measured in meters. The results of deferent measure for image-matching algorithms are shown in Table 1.

According results the deep learning-based matcher produces valid position estimates for 183 out of 200 images, corresponding to a localization success rate of 91.5%. The SIFT method produces 169 valid estimates, while ORB produces 155. Also, the mean positioning errors are 5.2 m for the deep learning-based approach, 7.5 m for SIFT, and 12.1 m for ORB. The corresponding RMSE values are 8.4 m, 11.5 m, and 19.4 m. In general, the deep learning-based matcher algorithm shows better results for all parameters**.**

Table 1. Results of comparison of image-matching algorithms.

Evaluation metric	**SIFT**	**ORB**	**Deep learning matcher**
Total test images	200	200	200
Valid position estimates	169	155	183
Localization success rate	84.5%	77.5%	91.5%
Mean error (m)	7.5	12.1	5.2
Median error (m)	4.7	7.7	3.2
RMSE (m)	11.5	19.4	8.4
90 percent error (m)	15.6	26.2	10.8

CONCLUSION

Determining a UAV's position by comparing camera images with maps is an important approach to visual localization. The method uses visual features shared by the captured image and a georeferenced reference map to establish the geographical location of the observed area.

The complete process includes image acquisition, preprocessing, feature extraction, descriptor matching, geometric verification, image registration, and coordinate transformation. Algorithms such as SIFT and ORB provide established methods for identifying and matching image features, while deep learning-based techniques offer additional possibilities for handling substantial differences in appearance and viewpoint.

The conversion from image coordinates to geographical coordinates requires an accurate reference map and a suitable geometric model. The estimated ground location must also be distinguished from the UAV's actual position, which depends on camera orientation, altitude, and terrain geometry.

Deep learning-based image-to-map matching provides a powerful framework for visual localization of drones. The proposed approach combines learned feature extraction, global image retrieval, local feature matching, attention-based correspondence estimation, and geometric verification.

The main advantage of this architecture is that it transforms the problem from direct pixel-level comparison into comparison of learned visual representations. This makes the system more robust to changes in viewpoint, scale, illumination, and image appearance. A hierarchical matching architecture is particularly suitable for large-scale applications. Global descriptors can first identify a small number of candidate

map regions, after which a local deep learning matcher performs detailed correspondence estimation. Finally, geometric verification can remove inconsistent matches and provide a reliable estimate of the drone's geographical position.

The resulting system can serve as the basis for GNSS-independent UAV localization and can be further extended by incorporating temporal information from consecutive drone frames, inertial measurements, digital elevation models, and multi-modal map information.

Although visual localization offers an alternative source of positioning information, its accuracy is affected by image quality, map resolution, environmental changes, and the reliability of feature correspondences. A robust implementation should therefore include geometric verification, performance evaluation, and methods for identifying uncertain or failed matches.

Further development of deep learning methods, multi-sensor fusion, and efficient image registration techniques may improve the reliability of visual positioning in complex environments.

Acknowledgement

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03 34/2026-03/200325)

References

[1] COUTURIER, A., AKHLOUFI, M. A.: A review on absolute visual localization for UAV. *Robotics and Autonomous Systems, 135*, 103666 (2021). https://doi.org/10.1016/j.robot.2020.103666

[2] FISCHLER, M. A., BOLLES, R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395,(1981). https://doi.org/10.1145/358669.358692

[3] https://github.com/IntelliSensing/UAV-VisLoc

[4] HU, S., CHANG, X.: Multi-view drone-based geo-localization via style and spatial alignment. arXiv, (2020).https://arxiv.org/abs/2006.13681

[5] MOSKALENKO, I., KORNILOVA, A., FERRER, G.: Visual place recognition for aerial imagery: A survey. *Robotics and Autonomous Systems, 183*, 104837, (2025). https://doi.org/10.1016/j.robot.2024.104837

[6] *BAYKAL, F. GEDIK, M.İ. REYES-ALDASORO, C.C. KARABAĞ, C.: Image Matching for UAV Geolocation: Classical and Deep Learning Approaches*. *Journal of Imaging, 11*(11), 409, (2025). https://doi.org/10.3390/jimaging11110409

[7] LOWE, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91–110, (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

[8] https://docs.opencv.org/4.13.0/da/df5/tutorial_py_sift_intro.html

[9] RUBLEE, E., RABAUD, V., KONOLIGE, K., BRADSKI,G.: ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2564–2571, (2011). https://doi.org/10.1109/ICCV.2011.6126544

[10] https://www.researchgate.net/figure/The-framework-of-Geometric-Correspondence-Oriented-FAST-and-Rotated-BRIEF-GC-ORB_fig1_337026817

[11] FU, Y., LUO, X., QIN, X., WAN, H., CUI, J., HUANG, Z.: Deep Learning-Based Feature Matching Algorithm for Multi-Beam and Side-Scan Images. *Remote Sensing*. 2025; 17(4):675. https://doi.org/10.3390/rs17040675

[12] ZHANG, S., LI, Z., ZHANG, K., LU, Y., DENG, Y., TANG, L., MA, J.: Deep learning reforms image matching: A survey and outlook. *arXiv preprint arXiv:2506.04619*, (2025).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

GENETIC ALGORITHM PARAMETER CONTROL FOR THE OPTIMAL DESIGN OF SPECTRALLY SELECTIVE 1D PHOTONIC CRYSTALS

ANJA PERIĆ
PhD student at Faculty of physics, University of Belgrade, Belgrade, anja.bartula994@gmail.com

OLGA JAKŠIĆ
Department of Microelectronic Technologies, Institute of Chemistry Technology and Metallurgy - National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia, olga.jaksic@ihtm.bg.ac.rs

Abstract: *This paper evaluates the impact of Genetic Algorithm (GA) parameter control and layer count on the optimization efficiency and performance of aperiodic 1D SiO2/TiO2 photonic crystals on glass. The target is a dual-band reflector at 1064 nm and 1550 nm. Calculated via the Transfer Matrix Method (TMM), robust GA parameters (Population 200, 300 Generations) present a bottleneck requiring 10,836.6 minutes (7.52 days) of CPU runtime. To resolve this, we analyze restricted GA parameters (Population 20, 10 Generations) across 2 to 17 pairs. Traditional quarter-wave stacks fail this dual-band requirement (fval = 0.2552). Under limited GA budgets, simpler designs yield poor reflectance: 2 pairs achieve R < 15% (fval = 0.0145), 5 pairs reach R ≈ 43% at 1064 nm and R ≈ 35% at 1550 nm (fval = 0.0807), and 14 pairs yield R ≈ 50% and R ≈ 62% (fval = 17.0666). High-performance spectra emerge only with larger stacks: 16 pairs achieve R ≈ 95% and R ≈ 80% (fval = 0.1918), while near-ideal response (R ≈ 98-99%) requires 17 pairs (fval = 0.0062). These findings establish a tuning framework to balance computational cost and optical design performance.*

Keywords: *photonic crystals, genetic algorithm, parameter control, transfer matrix method, dual-band reflector*

.

1. INTRODUCTION

Photonic crystals (PCs) are artificial nanostructured materials whose dielectric properties can be engineered to control the propagation of electromagnetic waves. One-dimensional (1D) PCs, formed as multilayer stacks of alternating high and low refractive index materials, have been widely utilized in applications such as infrared signature management, radiative cooling, and laser eye protection. In these practical contexts, the primary engineering task is to tailor the spectral characteristics - namely, absorptivity A(λ), reflectivity R(λ), and transmissivity T(λ) - across various electro-optical sensor bands. [1] A summary of key signature threats and sensor bands is shown in Table 1.

A particularly critical defense application is the dual-band high-reflection filter, capable of simultaneously attenuating reflection or providing laser shielding at key military laser wavelengths: the Nd:YAG laser operating at 1064 nm and the Er:Glass laser (often using Er-doped glass or solid-state materials) emitting at 1540 nm or 1550 nm.

Table 1. Signature threats and spectral regions in praxis

Region µm	Sensor threat	Practical example
0.7–1.1	NIR camera / laser range finder	Reflected ambient / laser light
1–2.5	SWIR imager	Reflected solar, thermal glow of hot objects
3–5	MWIR thermal imager	Thermal emission of warm objects (~300-800 K)
8–14	LWIR thermal imager	Th. emission, near-ambient objects (~250–350 K)

Table 2 outlines the characteristics of military lasers in praxis. Achieving high reflection (R > 99%) at these distinct, separated wavelengths while maintaining a neutral spectral response or high transmission in the visible and other infrared bands is a highly non-linear design problem. Traditional periodic dielectric mirrors, such as Bragg reflectors designed with quarter-wave optical thickness (QWOT), are intrinsically limited to a single fundamental bandgap and its higher-order harmonics.

Table 2. Military lasers in praxis

Laser	Wavelength	Use
Nd:YAG	1064 nm	Rangefinders, target designators
Nd:YAG 2nd harmonic	532 nm	Green laser dazzlers
Er:Glass	1540 nm	Eye-safe rangefinders
Ho:YAG	2090 nm	Some LIDAR systems
Diode	800-980 nm	Pumps, some direct-use systems

A QW stack optimized for 1064 nm cannot provide high reflectivity at 1550 nm, as it falls within a transmission window. Mixed QW designs, which combine multiple periodic stacks, suffer from destructive interference at the interface boundaries, shifting the bandgap edges and failing to yield a clean dual-band response.

Corresponding author: A. Perić;
E-mail: anja.bartula994@gmail.com
DOI: 10.5937/OTEH26139P

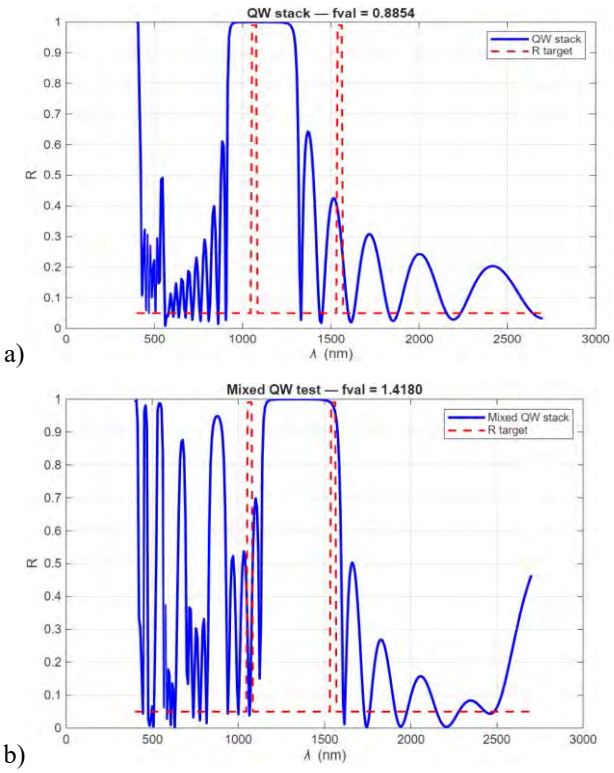


Figure 1. Spectral response comparison of traditional periodic quarter-wave (QW) stacks: a) A standard QW stack optimized for 1064 nm shows high reflection at the target but zero reflectance at 1550 nm; b) A Mixed QW stack shows severe band-edge shifting and interference oscillations in the background, resulting in a poor optimization score.

To break the symmetry constraints of periodic structures, aperiodic multilayer designs are required. In an aperiodic stack, the individual layer thicknesses are treated as independent variables, creating an extremely large, non-convex parameter space. Global optimization algorithms, such as Genetic Algorithms (GAs), are ideally suited for searching this complex landscape. However, direct optimization using GA is hindered by a severe, hidden computational bottleneck. Computing the electromagnetic response of a multi-layer stack using the Transfer Matrix Method (TMM) repeatedly across hundreds of generations and large populations is computationally prohibitive. For instance, executing a standard GA search with a population size of 200 over 300 generations requires an estimated 10,836.6 minutes (7.52 days) of continuous CPU runtime on standard platforms. This bottleneck forces engineers to utilize restricted GA options (e.g., small populations and limited generations), which risks premature convergence and poor spectral performance.

This paper addresses this challenge by systematically evaluating how the restriction and control of Genetic Algorithm parameters impact both the computational runtime and the optimized spectral profile. We analyze the performance of aperiodic SiO2/TiO2 stacks across different layer counts (from 2 to 17 pairs) and establish precise tuning guidelines to balance real-world computational budgets with high-performance optical requirements.

2. DESIGN METHODOLOGY AND TRANSFER MATRIX MODEL

The physical system under consideration consists of an interleaved, aperiodic stack of silicon dioxide (SiO2) and titanium dioxide (TiO2) layers deposited on a thick glass substrate (n = 1.5). The incident medium is air (n = 1.0). To ensure physical and industrial feasibility, the design is 'foundry-aware'. The overall thickness of any individual layer is strictly bounded between d_min = 20 nm and d_max = 400 nm, which complies with standard thin-film deposition limitations (such as magnetron sputtering or electron-beam evaporation).

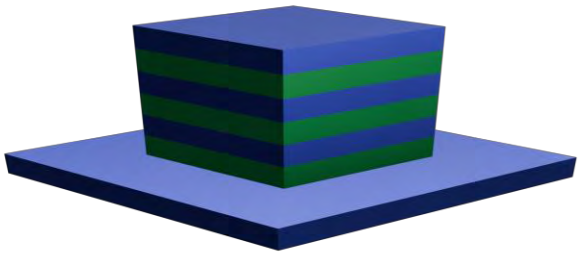


Figure 2. Schematic of the proposed 1D multilayer SiO2/TiO2 aperiodic stack on a glass substrate.

The electromagnetic response of the stack is modeled using the Transfer Matrix Method (TMM). For an N-layer system, the characteristic matrix of each layer is computed by relating the electric and magnetic fields at the boundaries. Refractive indices of SiO2 and TiO2 are dynamically evaluated to account for material dispersion across the target spectral range (400–2700 nm). TiO2 exhibits significant normal dispersion in the visible and near-infrared regions, which must be modeled accurately using experimental dispersion data to prevent model mismatch.

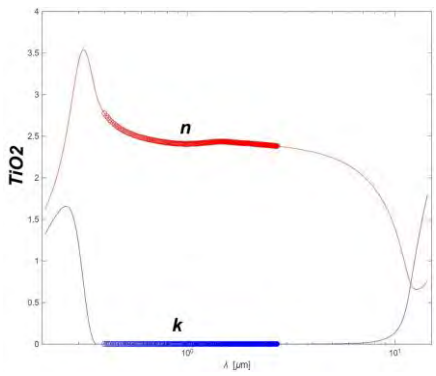


Figure 3. Refractive index (n) and extinction coefficient (k) dispersion of TiO2

To guide the optimizer toward achieving a dual-band high-reflection profile at 1064 nm and 1550 nm, while keeping reflectance low elsewhere, the objective function (fitness value, fval) is defined as a weighted mean-squared error between the calculated reflectance R_calc and the target reflectance R_target:

Fitness = Mean(W * (R_calc - R_target)^2)

The target reflectance profile R_target is set to 1.0 (100% reflection) in narrow bands centered around 1064 nm and 1550 nm, and 0.05 (5% reflection) elsewhere. Crucially,

the optimization weights (W) are heavily unbalanced: inside the two target laser bands, the weights are set to 100, while in the rest of the 400–2700 nm range, the weights are set to 1.

This spectral weight segmentation forces the genetic algorithm to prioritize satisfying the high-reflection criteria at the laser lines, preventing the optimizer from stalling in flat, non-critical spectral regions.

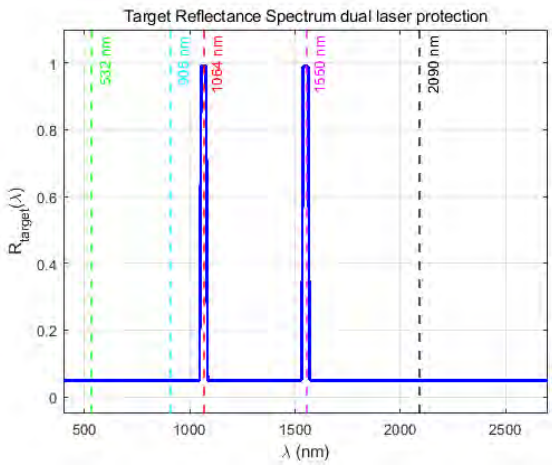


Figure 4. Target reflectance profile and corresponding band weights: The ideal target spectrum R_target showing narrow high-reflection peaks at 1064 nm and 1550 nm;

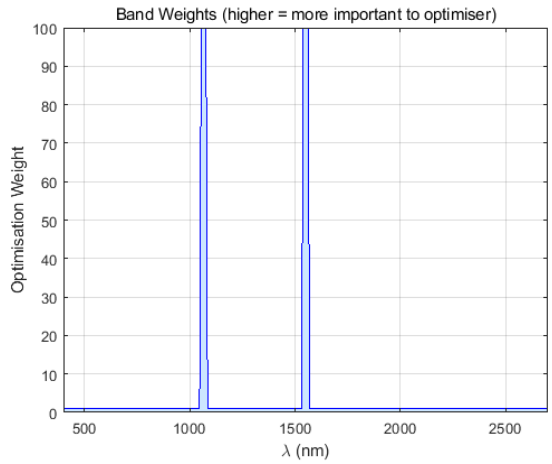


Figure 5. Target reflectance profile and corresponding band weights: The segmented optimization weights, highlighting the high-priority bands (Weight = 100) at the target laser wavelengths.

3. GENETIC ALGORITHM PARAMETER TUNING&CONVERGENCE ANALYSIS

The optimization was implemented in MATLAB using the global optimization toolbox. The search options are configured to explore the design space within the bound constraints [d_min, d_max] for N layers. The key parameters of the GA options block are configured as follows:

- **Population Size:** Controls the number of candidate designs in each generation. In comments of our original source, a size of 200 was desired but reduced to 20 to avoid extreme computational latency.
- **Max Generations:** The stopping criterion for the algorithm. A value of 300 was desired but reduced to 10 for rapid exploration.
- **Elite Count:** Set to 10. This ensures that the top 10 best-performing designs are copied directly to the next generation without modification, preventing the loss of high-quality designs.
- **Crossover Fraction:** Set to 0.8, meaning 80% of the next generation is produced by blending parent designs, maintaining a high rate of information recombination.
- **Mutation Function:** The adaptive feasible mutation function ('@mutationadaptfeasible') was chosen, generating small, random, self-adaptive changes to layer thicknesses while strictly respecting the 20 nm to 400 nm bounds.
- **Function Tolerance:** Set to 1e-8, defining the convergence threshold for termination.

To understand the optimization dynamics, a convergence run was analyzed for a stack with 10 pairs of layers (20 layers in total) over 100 generations. This analysis is critical for determining whether restricted GA parameters can yield stable, high-quality results.

The convergence behavior exhibits two distinct physical regimes. First, in the initial 0 to 15 generations, the mean fitness of the population drops exponentially from over 38.0 to less than 1.0. This represents the rapid elimination of highly unfit, random layer distributions. Second, the best fitness value (black dots) begins below 1.0 and exhibits a step-like stabilization, settling on a value of 0.338843 after 20 generations. This indicates that while the population is converging, the local search stagnates. This finding mathematically proves that restricting the algorithm to only 10 generations (as done in the quick-run mode) is premature. While the most drastic gains occur early, at least 20 to 50 generations are required for the best fitness value to mature and stabilize.

4. RESULTS AND DISCUSSION

We systematically evaluated the performance of optimized aperiodic SiO2/TiO2 filters across a range of layer pairs (N = 2 to 17 pairs, corresponding to 4 to 34 physical layers) under the restricted GA budget. The optimization results reveal a highly non-linear relationship between stack complexity, achieved minimum fitness values (fval), and spectral dual-band performance.

At low complexity, the structures are physically incapable of establishing a dual-band profile; with N = 2 pairs, the reflectance is flat and stays below 15% across the entire wavelength range (fval = 0.0145).

As the pair count increases to N = 5, distinct peaks emerge exactly at 1064 nm and 1550 nm, but under the restricted GA budget (PopulationSize = 20, MaxGenerations = 10), peak reflectance is limited to 43% at 1064 nm and 35% at 1550 nm (fval = 0.0807).

At N = 7 pairs, a severe optimization trade-off emerges: Case A yields a clean, quiet background (fval = 0.0792) but limits peaks to approximately 36%-37%; conversely, Case B achieves higher peaks of ~65% but at the cost of massive,

destructive oscillations (exceeding 90%) in the transmission bands (fval = 21.5104), violating the neutral background requirement.

Mid-complexity stacks with N = 14 pairs suffer from severe peak splitting and high noise across the spectrum (fval = 17.0666).

A critical performance threshold is crossed at N = 16 pairs, where reflectance rises to 95% at 1064 nm and 80% at 1550 nm (fval = 0.1918). Finally, with N = 17 pairs (34 layers) and a well-converged GA run, near-perfect dual-band reflection is achieved (fval = 0.0062), exceeding 98% and 99% reflectance respectively. However, as shown on the 17-pair spectrum, the resulting layer thicknesses are highly aperiodic and irregular (ranging from 48.02 nm to 397.06 nm). This spatial complexity is necessary to establish the localized cavity modes that trap and reflect the target laser lines.

However, manufacturing such a precise, 34-layer aperiodic stack with nanometer-scale tolerances represents a major physical fabrication challenge.

5. CONCLUSION

In this work, we successfully analyzed the impact of Genetic Algorithm (GA) parameter control and physical layer count on the optimization of aperiodic 1D multilayer SiO2/TiO2 photonic crystals. By systematically evaluating the trade-offs between computational budgets and optical performance, we draw several critical conclusions for the design of dual-band spectrally selective filters:

- **Parameter Optimization Balance:** While a full-scale GA search (Population Size = 200, Generations = 300) is computationally expensive, requiring over 7.5 days, a severe restriction (Population Size = 20, Generations = 10) leads to premature convergence, limiting peak reflectance to under 43%. Balanced configuration of Population Size = 50 and 50–100 Generations, runs in less than 24h while preventing local minima trapping.
- **Physical Complexity Threshold:** A minimum of 16 to 17 layer pairs (32 to 34 layers) is physically required to construct a dual-band bandgap with high reflectance (>95%) at both 1064 nm and 1550 nm. Low-complexity stacks (under 7 pairs) cannot provide dual-band selectivity regardless of the optimization duration.
- **Weight Segmentation Necessity:** The use of a heavily weighted objective function (W = 100 at target lines, W = 1 elsewhere) is absolutely critical. It prevents the genetic algorithm from stagnating and guides the search toward establishing high-reflection bands in highly localized spectral regions.

Future work will focus on integrating experimental fabrication tolerances directly into the GA mutation function and exploring the optical performance of these optimized aperiodic stacks under oblique angles of incidence.

Acknowledgements

This research has been financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-33/2026-03/200026). This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7451, Plasmonic-based light harvesting for photocatalytic microfluidic devices - PlasmaHarvest.

References

[1] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, 'Deep learning for the design of photonic structures', *Nat. Photonics*, vol. 15, no. 2, pp. 77–90, Feb. 2021, doi: 10.1038/s41566-020-0685-y.

[2] P. Vincent *et al.*, 'Application of Genetic Algorithm for More Efficient Multi-Layer Thickness Optimization in Solar Cells', *Energies*, vol. 13, no. 7, p. 1726, Apr. 2020, doi: 10.3390/en13071726.

[3] J. D. Joannopoulos, J. N. Winn, and S. G. Johnson, *Photonic Crystals: Molding the Flow of Light - Second Edition*. Princeton, NJ: Princeton University Press, 2011. doi: 10.1515/9781400828241.

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION VI

Telecommunication and information systems - TIS

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AN EXAMPLE OF MISSILE-TO-TARGET INTERCEPTION

ZVONKO RADOSAVLJEVIĆ
Millitary technical institute Belgrade, zvonko.radosavljevic@mod.gov.rs

DRAGAN NIKOLIĆ
Millitary technical institute Belgrade, nikolicdragansiki@gmail.com

DARKO PIJEVIĆ
Millitary technical institute Belgrade, darko.pijevic@gmail.com

DEJAN IVKOVIĆ
Millitary technical institute Belgrade, divkovic555@gmail.com

Abstract: *Methods for calculating the interception point of the trajectory of a projectile and a moving target are the subject of many projects and are being solved in several directions. Initially, these were trigonometric methods, while nowadays probabilistic methods or methods of integrating differential equations are used. The paper proposes a classical method of integrating differential equations of projectile motion using the Runge-Kutta method, based on the initial azimuth and elevation angles of the projectile launch parameters, using the known initial projectile velocity. Future target positions are predicted based on the radar target tracking and calculated meeting time of the projectile and the target, using the interpolation method of the target's position coordinates and acceleration. The simulation results showed satisfactory accuracy for targets in the air at a distance of 2000 m with standard deviation of error position on 15 m.*

Keywords: *Target tracking, Kalman filter, Taylor extrapolation polynom, ballistic trajectory.*

1. INTRODUCTION

Intercepting an incoming artillery or mortar round is receiving interest as an area-protection mechanism. A problem has been that estimation, based on 6-degrees of freedom -(DoF) equations for both the incoming round and the interceptor round or missile, is to time consuming from a computational perspective. This calculation is used to find the maneuver control and the aimpoint or lead angle needed to intercept an incoming round. The method uses reduced dynamics flight models for the interceptor and the incoming round. Also, questions relating to the probability of a hit given a specific standard deviation and bias are addressed [1].

Guidance laws that aim to achieve desired impact angle, as well as zero miss distance, are pushed forward by a number of techniques [2]. Some authors derived guidance laws with impact angle constraint based on proportional navigation guidance (PNG). These PNG-based impact angle control guidance techniques are simple in form and cost little in the calculation, but some constraints on the impact angle cannot strictly be satisfied when the initial heading coincides with the line-of-sight (LOS). The sliding mode control (SMC) technique is another commonly used technique in the field of impact angle control guidance (IACG). Other authors use SMC to achieve desired impact angle. Guidance laws based on a conventional SMC can only ensure asymptotic convergence. A guidance law based on terminal sliding mode control was proposed in [3], which satisfied the finite-time convergence. So, it is also posible for combined SMC and virtual target strategy to control collision time and angle. Guidance laws based on SMC have strong robustness, which are even effective on maneuvering targets. However, the discrete terms in the guidance formula always cause inevitable oscillations in the guidance commands, which is not suitable for actual implementations. The SMC technique was introduced to design a guidance law that makes multiple missiles hit the target in a preset direction at the same time [4], and it utilized range-to-go and radial relative velocity rather than time-to-go to construct the sliding surface to avoid performance degradation. Some authors combined SMC, backstepping, and super-twist extend state observer to propose a robust guidance law [5] and verified its performance by a dynamics model.

This paper is aimed at proposing a novel algorithm, which calculates firing parameters (cannon barrel azimuth and elevation) only the data obatined from from radar sensor. When the target enters the projectile's zone of action, an algorithm is activated that, in the first step, extrapolates the future position of the target in the time interval required for the projectile to reach the target. Given that the projectile's speed is an order of magnitude greater than the target's speed of movement, the integration of the projectile's differential equations of motion calculates the azimuth and elevation angle of the gun barrel in order to meet the target, under

Corresponding author: Z.Radosavljevic;
E-mail: zvonko.radosavljevic@mod.gov.rs

DOI: 10.5937/OTEH26077R

the condition of a minimum distance between the target and the projectile.

The paper is organized as follows: after introductory considerations, theoretical models of the target and projectile are given. The third chapter describes the proposed algorithm for predicting the collision of the target and projectile, while the fourth chapter is dedicated to presenting the results of numerical simulations. Finally, concluding considerations are given.

2. PROBLEM STATEMENTS

2.1 Targets and measurements models

Consider a discrete-time stochastic hybrid system:

$$x_k = f(x_{k-1}) + q_{k-1} \tag{1}$$

$$y_k = h(x_k) + r_k \tag{2}$$

where $x_k \in R^n$ is the state vector in effect at sample time k with dynamic model function $f(.)$ and transition matrix F_k, $y_k \in R^m$ is the vector of observations in effect at time k, $q_{k-1} \approx N(0, Q_{k-1})$ is the Gaussian process noise, $r_k \approx N(0, R_k)$ is the Gaussian measurements noise.

2.2 Missile dynamic models

Consider a 3D motion scenario for a missile against a moving target. Inertial, moving and bound coordinate systems are used to model the movement of the target and projectile (Figure 1).

Positions are x, y and z and velocities are v_x,v_y and v_z and θ and φ their azimuth and elevation flight path angles, respectively. State vector is given by the $X = [x\ v_x\ y\ v_y\ z\ v_z]$. R stands for the distance between the missile and the target. The engagement kinematics for this problem can be given as follows [8]:

$$\begin{aligned} x &= x_0 + v_{x0} t \\ y &= y_0 + v_{y0} t \\ z &= z_0 + v_{z0} t - g\frac{t^2}{2} \end{aligned} \tag{3}$$

$$\begin{aligned} v_{x0} &= v_0 \cos\varphi \cos\theta \\ v_{y0} &= v_0 \cos\varphi \sin\theta \\ v_{z0} &= v_0 \sin\varphi \end{aligned} \tag{4}$$

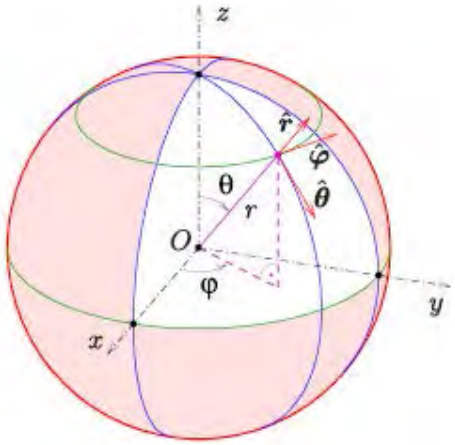


Figure 1. Kugel model

For ease of calculation, the problem can be reduced to 2D dimension (x,y). Due to the action of gravity, at time t the vertical component of the initial velocity decreases by the value of gravity, while the horizontal component remains unchanged. At that position, the velocity vector v makes an angle with the horizontal. The differential equations of motion in this case are [9]:

$$\begin{aligned} m\ddot{x} &= 0 \\ m\ddot{y} &= -mg \end{aligned} \tag{5}$$

or

$$\begin{aligned} \ddot{x} &= 0 \\ \ddot{y} &= -g \end{aligned} \tag{6}$$

The general solution of this system of equations is:

$$\begin{aligned} x &= C_1 t + C_2 \\ y &= -\frac{1}{2} g t^2 + C_3 t + C_4 \end{aligned} \tag{7}$$

The integration constants are determined by the initial conditions at the time:

$$\begin{aligned} x_0 &= 0, \\ y_0 &= 0, \\ (\dot{x})_0 &= v_0 \cos\alpha_0, \\ (\dot{y})_0 &= v_0 \sin\alpha_0, \end{aligned} \tag{8}$$

The final equations of motion of the projectile now are:

$$\begin{aligned} x &= v_0 t \cos\alpha_0 \\ y &= -\frac{1}{2} g t^2 + v_0 t \sin\alpha_0 \end{aligned} \tag{9}$$

By eliminating t from these parametric equations, the equations of the trajectory are obtained

$$y = -\frac{g}{2 v_0^2 \cos^2\alpha_0} x^2 + tg\alpha_0 x \tag{10}$$

This is the parabola that lies symmetrically with respect to

the coordinates of the vertex of the trajectory, i.e. parabola with a vertical axis and a concave side downward. If it is assumed that $0<\alpha_0<\pi/2$ the value is positive and according to (5) the abscissa x increases with time i, i.e. the projectile is constantly moving to the right. To determine the range l, expresses the range of the projectile:

$$l=\frac{v_0^2}{g}\sin(2\alpha_0) \tag{11}$$

The total flight time t_A of the projectile from 0 to A is obtained by the equation:

$$t_A=\frac{l}{v_0\cos\alpha_0}=\frac{v_0{}^2\sin(2\alpha_0)}{gv_0\cos\alpha_0}=\frac{2v_0}{g}\sin\alpha_0 \tag{12}$$

The projections are twice the values for the abscissa of the vertex of the path and the duration of the flight to it, which must also be for reasons of symmetry. The height of the vertex of the path is obtained by the:

$$h=-\frac{g}{2}t_T^2+v_0\sin\alpha_0 t_T=\frac{v_0^2}{2g}\sin^2\alpha_0 \tag{13}$$

where t_T is time necessary to rich h.

3. DESCRIPTION OF PROPOSED ALGORITHM

3.1 Taylor polynomials

Definition.
Let f be a function that has all derivatives up to order n at a. Then we define the **Taylor polynomial** of f of degree n with center a as [10]:

$$T_n(x)=f(a)+f'(a)(x-a)+\frac{1}{2!}f''(a)(x-a)^2+\ldots+\frac{1}{n}=\sum_{k=0}^{n}\frac{f^{(k)}(a)}{k!}(x-a)^2 \tag{14}$$

Theorem (Taylor's theorem).

Consider two distinct real numbers a and x. Let I be the closed interval with these two points as endpoints. Assume that a function f has continuous derivatives up to order n on I and $f^{(n+1)}$ on the interior Int(I). Then

$$Rn(x)=\int_a^x\frac{f^{(n+1)}(t)}{n!}(x-t)^n\,dt \tag{15}$$

Moreover, there exists a number c between a and x such that

$$Rn(x)=\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{16}$$

The first result is called the **integral form** of the remainder, the second is called the **Lagrange form**. In fact, these precise results are also not exactly useful, since the integral may easily be too difficult to evaluate, while the second statement is existential, we know that such a c exists, but we have no idea what it is. However, the Lagrange form can be approximated from above and with a bit of luck it turns out that the upper estimate is small.

3.2 Projectile-target interception

Basic idea was devised to indicate the time needed to hit an incoming target, with a defined angle of hitting the target. The incoming round was modeled using the point interpolation equations and the interceptor round was modeled with a single dimension differential equation. Using these equations, it was possible to compute the lead angle needed for the gun interceptor round. This calculation is based on estimated time of arrival of the interceptor at predicted points of the target trajectory.

The expected time of the interceptor's arrival at each point can be calculated. At each point, the predicted time until arrival of the target and the interceptor are differenced. An aim point/lead angle can be selected for any positive difference by adjusting the firing time of the interceptor so that the time difference is zero. It is also possible to search the interval that goes from negative to positive for the zero delay aim point. The terminal geometry could also be considered in selecting an aim point. Post launch adjustments are made by selecting a different predicted target trajectory point for the interceptor to hit. After the interceptor is fired, it could only adjust its course by one degree at a time; thus, corrections could only take place if there was over one degree of error in the predicted impact point. The accuracy of the target track would have to meet certain criteria before an interceptor could be fired. Meeting this accuracy requirement would ensure that the target would be within the interceptor's maneuver space. Likewise the estimate of the interceptor's trajectory is of equal concern. As time increases, the accuracy of both the estimate of the target trajectory and the interceptor trajectory will increase. Many times, polynomial filters are discussed as candidates for trajectory estimation; these filters do not attempt to capture the dynamics of the situation and for longer prediction times are inadequate. Although filters based on *6-DoF* models are conceivable, they impose a computational load that cannot be met at this time [11].

3.3 Steps of proposed method

Let's assume the *3D* movement of a target in the air (e.g. a drone or quad copter), which moves in a straight line and with a slight maneuver (not larger than 3g). The target is tracked during the experiment by a Kalman filter (KF) using a CV (constant velocity) motion model. At the moment the target enters the weapon's engagement zone, an automatic 'encounter' calculation is activated. It is basically a three-dimensional problem of the kinematics

of the movement of the projectile and the target. It is necessary to solve the pure pursuit ballistic problem and its 'encounter' with the target that has entered the engagement zone of the weapon. For this calculation, the known laws of motion of a projectile fired from a known location, with a known initial velocity, are used. The algorithm should determine the angle at which the projectile is fired. The algorithm is performed in five main steps. These are the steps of the algorithm:

1. The target is tracked using a radar sensor in the C band, moving the state estimator (*Kalman filter*)

2. Calculation of the target distance from the weapon

3. Extrapolation by Taylor polynomial (most often of the third degree) of the future state of the target for the time it takes for the projectile to cover the calculated distance to the target. The order of the Taylor polynomial depends on the target trajectory until the moment of entering the zone of action. Based on the azimuth and elevation angles, the weapon is positioned towards the target

4. Integration of the differential equation of motion of the projectile by the *Runge-Kutta* method, using the ode 45 function. the parameter (i.e. condition) that needs to be met to complete the integration is the distance of the projectile to the target, which should be below 10m

5. Firing a sequentially with different (close) angles, so that the proximity fuses on the projectiles are activated

4. RESULTS OF SIMULATION

For experimental purpose we use the two type of scenario:

A) Simulated scenario with constant velocity of target (Fig.3)

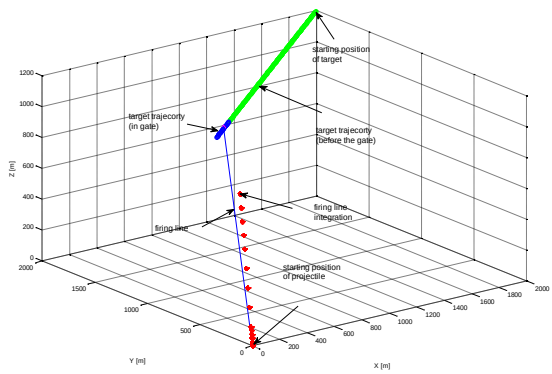


Figure 3. Simulated scenario with constant velocity

Using the known initial conditions, an expected projectile path can be generated. Range and velocity measurements can be used to improve the estimate of the projectile's path. The differential equation and the measurement process are combined to form a Kalman filter. The differential equation models the physics of the trajectory while the measurement updates the parameters of the equation based on least squares estimation.

$$F_{cv} = \begin{bmatrix} 1 & dt & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & dt & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (17)$$

The simulation has 500 time intervals of duration dt=0.1[s] from the radar. The target state vector has $x=[x\ v_x\ y\ v_y\ z\ v_z\ ]$. The initial target parameters are: $x_t= 2000$, $v_{xt} = -80$, $y_t = 2000$, $v_{yt} = -60$, $z_t = 1200$, $v_{zt} = -5$. The cannon is launched at the coordinate origin. The target moves from the initial position, towards to the zone of action and when it enters the zone of action, the differential equation of motion of the projectile is integrated until the time of flight of the projectile until the encounter with the target is obtained.

B) Real data scenario (Fig.4).

These times are given in Table 1 for both scenarios. The distance of the cannon's effective range is 2000 m. The initial velocity of the projectile is 950 m/s. The target continues to move until the condition is met that it meets the projectile within the calculated time, in some delta environment. Transition matrix of constant acceleration (CA) target dynamic motion is:

$$F_{ca} = \begin{bmatrix} 1 & dt & dt^2/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & dt & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & dt & dt^2/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & dt & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & dt & dt^2/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (18)$$

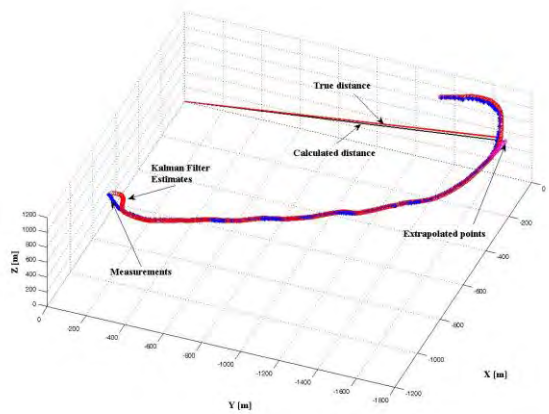


Figure 4. Simulation with real radar data

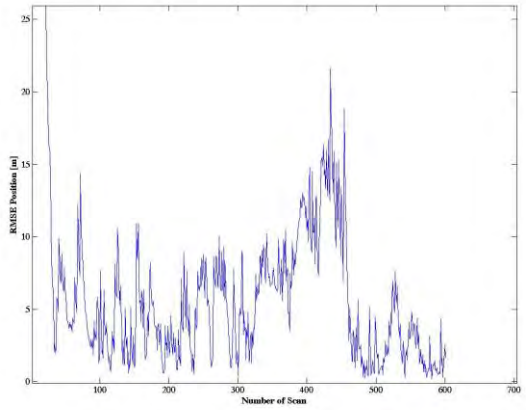


Figure 5. RMSE position diagram for real data scenario

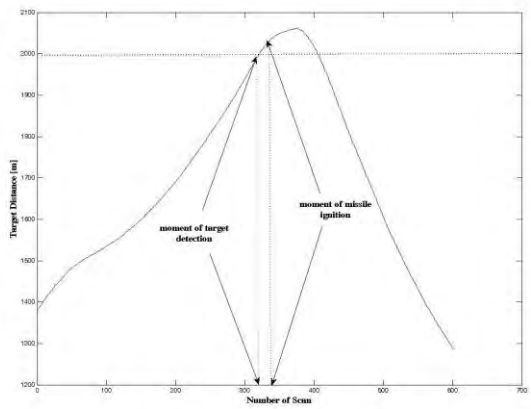


Figure 6. Distance of target diagram

The experimental results are given in the form of diagrams of the mean square error per position (real data – state estimate when tracking the target with the Kalman filter) given in Fig. 5. Figure 6. shows a diagram of the distance between the gun position (the initial position of the projectile) and a moving target. When the target enters the projectile's zone of action, an algorithm is activated that, in the first step, extrapolates the future position of the target in the time interval required for the projectile to reach the target. Given that the projectile's speed is an order of magnitude greater than the target's speed of movement, the integration of the projectile's differential equations of motion calculates the azimuth and elevation angle of the gun barrel in order to meet the target, under the condition of a minimum distance between the target and the projectile.

5. CONCLUSION

A complete algorithm for detecting, tracking and engaging low-flying targets using radar band range was proposed and tested in this paper. The algorithm showed good results in engaging moving targets in the air, using the Taylor polynomial (in the domain of extrapolation of future target states) and in the domain of predicting the missile and target collisions, using the Kalman filter. Preliminary research results are presented. The errors in approaching the bullet and targets are between 10 and 15 [m]. In future, a functional model for the complete detection, tracking and engaging low-flying targets in the air is envisaged.

References

[1] N. A. Shneydor, Missile Guidance and Pursuit: Kinematics, Dynamics and Control. Elsevier, Jan. 1998.

[2] Z. Radosavljevic, D. Ivkovic and B. Kovacevic: The Prediction of Missile-Target Interception Using C-band Radar, 13th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Silver lake, Republic of Serbia, 8-11 June 2026.

[3] I. S. Jeon, J. I. Lee, and M. J. Tahk, “Impact-time-control guidance law for anti-ship missiles,” IEEE Transactions on Control Systems Technology, vol. 14, pp. 260–266, Mar. 2006.

[4] M. J. Tahk, C. K. Ryoo, and H. Cho, “Recursive time-to-go estimation for homing guidance missiles,” IEEE Transactions on Aerospace and Electronic Systems, vol. 38, pp. 13–24, Jan. 2002.

[5] H. S. Shin, H. Cho, and A. Tsourdos, “Time-to-go Estimation using Guidance Command History,” IFAC Proceedings Volumes, vol. 44, pp. 5531–5536, Jan. 2011.

[6] N. Dhananjay and D. Ghose, “Accurate Time-to-Go Estimation for Proportional Navigation Guidance,” Journal of Guidance, Control, and Dynamics, vol. 37, pp. 1378–1383, July 2014.

[7] S. R. Kumar and D. Ghose, “Sliding mode control based guidance law with impact time constraints,” in 2013 American Control Conference, (Washington, DC), pp. 5760–5765, IEEE, June 2013.

[8] R. Tekin, K. S. Erer, and F. Holzapfel, “Polynomial Shaping of the Look Angle for Impact-Time Control,” Journal of Guidance, Control, and Dynamics, vol. 40, pp. 2668–2673, Oct. 2017.

[9] R. Tsalik and T. Shima, “Circular Impact-Time Guidance,” Journal of Guidance, Control, and Dynamics, vol. 42, pp. 1836–1847, Aug. 2019.

[10] J. I. Lee, I. S. Jeon, and M. J. Tahk, “Guidance law to control impact time and angle,” IEEE Transactions on Aerospace and Electronic Systems, vol. 43, pp. 301–310, Jan. 2007.

[11] T. H. Kim, C. H. Lee, I. S. Jeon, and M. J. Tahk, “Augmented Polynomial Guidance With Impact Time and Angle Constraints,” IEEE Transactions on Aerospace and Electronic Systems, vol. 49, pp. 2806–2817, Oct. 2013.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DEEPFAKE TECHNOLOGY AND THE TRANSFORMATION OF TRUST IN CONTEMPORARY INFORMATION SYSTEMS

SANELA VELJKOVIĆ
Vinča Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, Department of Physical Chemistry, Belgrade, Serbia, sanela.veljkovic@vin.bg.ac.rs

MARINA DABETIĆ
Vinča Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, Department of Physical Chemistry, Belgrade, Serbia, fmarina@vin.bg.ac.rs

ANITA KLIKOVAC
Ministry of Internal Affairs, Emergency management sector, Smederevo, Serbia, anitaklikovac@yahoo.com

Abstract: *In recent years, there has been a rapid development of deepfake technology, and it has become increasingly widespread. It appears that there are virtually no limitations to this technology, which has advanced to the point where anyone's likeness or voice can now be replicated. Distinguishing between authentic and artificially generated multimedia content is becoming increasingly difficult, and an average person can hardly detect discrepancies. The misuse of deepfake technology can occur in various ways – including the creation of pornographic content, the dissemination of fake news, and the execution of different types of fraud. In contemporary conflicts, fabricated audio and video recordings can provoke panic and influence decision-making processes. Deepfake technology has already been used as a supporting instrument of warfare, serving to spread panic and destabilize societies. The likenesses of current political leaders have been exploited to create fabricated content. The implications of such developments may be significant, and it appears that a transformation of trust in contemporary information systems has occurred. Therefore, the aim of this paper is to analyze the development of deepfake technology and to examine the role that the misuse of this technology – including the manipulation of multimedia content – plays in contemporary conditions.*

Keywords: *deepfake, contemporary information systems, trust, fabricated content*

1. INTRODUCTION

Information technologies have developed to such an extent that today it seems as though almost anything is possible. The widespread availability of the Internet, along with various artificial intelligence tools, has significantly facilitated everyday life. Nevertheless, despite the numerous benefits associated with technological progress, there are also certain risks and negative consequences. Deepfake technology first emerged several years ago, primarily in the form of applications designed for entertainment purposes. Through these applications, users could place their own or another person's face onto the body of someone else, or create videos in which individuals appeared to say things they had never actually said. During the early stages of its development, deepfake technology was still relatively limited, making it easy to recognize such content as artificial and unrealistic. Today, however, the line between authentic and fabricated content has become increasingly blurred, leaving individuals unable to place complete trust in what they see with their own eyes. In addition to its many positive applications, the development and refinement of deepfake technology have also generated numerous concerns. The growing difficulty of distinguishing between authentic and manipulated content has resulted in various forms of misuse. It is estimated that the consequences of deepfake technology extend beyond individual victims and may also pose threats to national security [1, p. 1443]. Detecting deepfake content has become increasingly challenging because the technology has advanced considerably, while detection methods remain complex and often inaccessible to the average user. What began as a technology intended primarily for entertainment now possesses the potential to "create fake news within communities, spread hate and disinformation that manipulate public opinion, and facilitate malicious activities such as extortion, blackmail, spoofing, identity theft, reputational damage, and deepfake pornography" [2, pp. 143296–143297]. In the context of pornography, celebrities, singers, and actresses were once the primary targets of deepfake content. Today, however, due to the widespread use of social media and the rapid advancement of deepfake technology, virtually any woman can become a target. Similarly, numerous fraud cases involving the use of deepfake technology have already been reported worldwide. As the technology continues to improve, distinguishing between genuine and fabricated content becomes increasingly difficult. Moreover, "if deepfakes or synthetic media are weaponized, trust in digital content may erode, making it difficult for individuals to rely on media even when the

Corresponding author: S. Veljković;
E-mail: sanela.veljkovic@vin.bg.ac.rs
DOI: 10.5937/OTEH26078V

content is authentic" [3, p. 243]. Consequently, the aim of this paper is to analyze the development of deepfake technology and examine the role that its misuse plays in contemporary conditions. The first section of the paper presents the development of deepfake technology and the methods used to detect deepfake content. The second section focuses on the transformation of trust in contemporary societies and the broader consequences associated with the use of this technology. Finally, the paper concludes with a summary of the findings and provides a list of references used in its preparation.

2. PRODUCTION AND DETECTION OF DEEPFAKE CONTENT

A significant number of deepfake videos can be viewed as harmless entertainment or artistic works that make use of the likenesses of public figures, whether living or deceased [4, p. 43]. However, in many cases, this technology is not used solely for positive purposes. Consequently, Nasiri and Hashemzadeh raise a logical question: "Should artificial intelligence developers bear some degree of responsibility for the actions of malicious actors who misuse their products?" [3, p. 244]. The first widely known deepfake content was a pornographic video featuring a celebrity that appeared on Reddit in 2017. Shortly thereafter, applications such as FakeApp and FaceSwap emerged, while the greatest concern was generated by the application DeepNude, which became known as a "digital undressing application" [5, p. 607]. Deepfake technology enables the creation of images and videos depicting events that never actually occurred. As a result, the face or voice of virtually any individual can be exploited to generate fabricated content. Deepfake refers to "an artificial intelligence-based technology capable of manipulating images, audio recordings, and videos to depict an event that never actually happened" [6, pp. 178–179]. More specifically, "the term 'deepfake' originates from the underlying deep learning (DL) technology, which enables face swapping in digital content and creates the false impression of an individual in a realistic environment" [2, p. 143297]. Such content is produced through various techniques, including autoencoders and Generative Adversarial Networks (GANs) [5,6]. As this technology continues to advance, the differences between genuine and manipulated content, which were once relatively easy to detect, have become increasingly difficult to identify. Consequently, "these manipulated media can portray individuals saying or doing things they never actually said or did, potentially leading to serious consequences in political, legal, and social contexts" [7, p. 73]. Owing to the widespread use of social media and the extensive sharing of photographs and videos, individuals today are more exposed than ever before. Likewise, the likenesses and voices of public figures - including actors, singers, and politicians - can be exploited to create deceptive and misleading content.

As Sophia observes, "early deepfake models primarily focused on manipulating video and audio content, whereas more recent advances have enabled artificial intelligence to generate highly convincing synthetic media in the form of text, images, and even entire virtual personas" [7, p. 83]. In addition to the creation and distribution of deepfake pornographic material, such content represents "a serious threat to society, political systems, and the business sector, as it places pressure on journalists who struggle to distinguish truthful information from falsehoods, threatens national security through the dissemination of propaganda capable of influencing elections, undermines citizens' trust in information provided by authorities, and raises significant cybersecurity concerns for both individuals and organizations" [4, p. 47]. Deepfake technology has already been used to create videos depicting political leaders expressing opinions and making statements that they would be highly unlikely to utter in reality. Furthermore, particularly during election campaigns, this technology enables candidates to "appear more attractive, younger, better-looking, and more energetic; communicate with multiple audiences simultaneously in different languages; and personalize their messages for individual voters" [8, p. 468]. As a result, the distinction between truth and falsehood becomes increasingly blurred, while public trust and confidence in institutions may be seriously undermined. Due to the rapid advancement of this technology, individuals can no longer be certain whether a video recording or photograph is authentic or artificially generated. Trust is an essential prerequisite for the effective functioning of institutions. Without it, "when disinformation undermines institutions, their ability to implement policies, enforce laws, and manage public crises is severely compromised" [7, p. 78]. Therefore, contemporary information systems must be capable of providing citizens with relevant, accurate, and reliable information, while also detecting deepfake content in a timely manner and preventing its further dissemination.

Methods for detecting deepfake content vary considerably, while the technology itself has advanced so rapidly that ordinary individuals can no longer confidently trust their own eyes. Whereas early deepfake videos often contained visible flaws that could be identified by human observation, this is no longer the case. Regarding detection techniques, Cinar identifies two primary approaches: the analysis of biological signals and the detection of pixel-level irregularities [6, pp. 182–183]. The detection of manipulated content has become a major challenge. As Koopman et al. note, "it is increasingly likely that everyday video evidence has been manipulated, creating a growing need for verified authentication methods capable of detecting deepfake alterations" [9, p. 133]. In response to this challenge, Westerlund identifies four principal approaches to combating deepfake content: legislation and regulatory measures; corporate policies and voluntary initiatives by private companies; education and training; and technological countermeasures, including deepfake detection, content authentication, and the prevention of deepfake creation [4, p. 44]. States are increasingly classifying deepfake content as a potential security threat and are seeking to establish appropriate legal frameworks to address this issue. In this regard, the European Union has adopted the Digital Services Act (DSA) and the Artificial Intelligence Act (AIA), both of which aim to regulate harmful and deceptive online

content. In the United States, various federal and state-level initiatives have been introduced to address the challenges posed by deepfakes, including the "Deepfake Accountability Act." A similar initiative exists in China through the "Regulations on the Administration of Deep Synthesis of Internet Information Services." In addition, major technology companies and online platforms, such as Meta and Google, have recognized the need for greater oversight of artificially generated content [7, p. 81]. Nevertheless, the transformation of trust in contemporary information systems has already occurred, to a considerable extent, as a consequence of the widespread development and misuse of deepfake technology.

3. TRANSFORMATION OF TRUST IN CONTEMPORARY SOCIETY

The inability to identify flaws when viewing manipulated content raises an important question: how can individuals distinguish between authentic and fabricated information? The use of deepfake technology "raises questions about human perception and challenges the ability to trust one's own eyes and ears when watching a video of a politician discussing important international, regional, and national issues" [1, p. 1449]. As a result, trust in contemporary information systems, political structures, and public institutions may be directly undermined, potentially leading to various political and social crises. In many countries, the use of deepfake technology remains insufficiently regulated by law. Consequently, individuals who create and distribute such content often face little or no legal accountability for their actions. Some scholars characterize the present era as an age of post-truth, in which the distinction between reality and fiction becomes increasingly difficult to establish. In this context, they warn that "when audiences lose trust in established media organizations, they may turn to alternative sources that prioritize sensationalism over accuracy, thereby further deepening the disinformation crisis" [7, p. 77]. The spread of disinformation and fake news has been significantly facilitated in recent years by deepfake technology and artificial intelligence tools. Consequently, "the manipulation of public discourse through AI-driven disinformation campaigns can shift political priorities, complicate legislative processes, and create an environment in which evidence-based policymaking becomes increasingly difficult" [7, p. 77]. Numerous political leaders have become targets of deepfake content, while others have reportedly used this technology themselves for propaganda purposes [1]. Although deepfake technology was initially developed primarily for entertainment and recreational purposes, its continuous advancement has led individuals, organizations, and governments to reconsider the risks associated with its use. Within only a few years of its emergence, deepfake technology has demonstrated a wide range of harmful applications, prompting growing efforts to place its use under stricter oversight and to establish comprehensive legal regulations governing its development and deployment.

As previously noted, many celebrities, singers, and actresses have become victims of the misuse of deepfake technology through the creation of fabricated pornographic content. Furthermore, numerous women were placed at risk following the emergence of the so-called "digital undressing application." Similarly, many politicians have been confronted with highly realistic videos depicting them making statements that they never actually uttered. Such videos can spread rapidly through social media and various online platforms. By the time it is established that a recording is fabricated or generated through deepfake technology, significant damage may already have occurred. In 2018, a deepfake video created by a Belgian political party was released with the aim of encouraging citizens to sign a petition. In the video, Donald Trump appeared to advise Belgian citizens on climate change, provoking considerable public outrage [4, p. 43]. In the same year, an attempted military coup took place in Central Africa following the release of a video featuring the President of Gabon, Ali Bongo Ondimba. At the time, many citizens believed that the president was either seriously ill or had already died due to his prolonged absence from public life, which fueled suspicions regarding the authenticity of the recording [4, p. 44]. This case serves as an illustrative example of the consequences that may arise when the public loses confidence in the authenticity of video evidence. Deepfake technology was also employed during the COVID-19 pandemic, when "deepfake videos and audio recordings circulated online, often falsely attributing statements to prominent public health officials, politicians, or celebrities" [3, p. 237]. During a global health crisis of this magnitude, individuals not only faced fears concerning their own health and the well-being of their loved ones but were also confronted with an overwhelming volume of both accurate and false information circulating simultaneously. The coexistence of reliable information and fabricated content further complicated public understanding of the crisis and contributed to the spread of uncertainty and distrust.

The use of deepfake technology during election campaigns is not a recent phenomenon. As Nasiri and Hashemzadeh observe, "while traditional election interference techniques involved the dissemination of misleading information through text-based fake news, the 2020 elections were marked by the increasing use of deepfakes and AI-generated content to manipulate voters, undermine trust in institutions, and influence electoral outcomes" [3, p. 238]. An increasing number of scholars have examined the impact of deepfake content in the electoral context. For example, Łabuz and Nehring analyzed the use of deepfake technology in eleven countries - the United States, Turkey, Argentina, Poland, the United Kingdom, France, India, Bulgaria, Taiwan, Indonesia, and Slovakia - during 2023, with the aim of identifying its potential implications for future electoral processes. Their conclusion is particularly noteworthy: "we believe that none of the cases in which deepfake content has thus far been used in the context of elections has had a decisive impact on the outcome of the elections themselves. This does not mean, however, that there has been no effect. (...) Although dystopian predictions of an 'information apocalypse' have not (yet?) materialized, there are already visible signs of declining trust in

information, politics, and the media, which reinforce feelings of uncertainty within society and create new opportunities for manipulation" [8, p. 468]. These findings suggest that deepfake technology can have a significant impact on public trust in information systems, even when it does not directly alter election results. Nevertheless, the use of deepfake technology is not limited to peacetime conditions. On the contrary, it has increasingly been employed during armed conflicts. According to Twomey et al., "the Russia–Ukraine war represents the first real-world example of the use of deepfake technology in warfare, with a number of incidents involving deepfake content featuring Russian and Ukrainian government officials being used for purposes of disinformation and entertainment" [10, p. 1]. In this context, several notable examples emerged, including a fabricated video of a Ukrainian fighter aircraft, a deepfake video depicting Vladimir Putin announcing peace with Ukraine, and a deepfake recording of Volodymyr Zelenskyy purportedly surrendering to Russia [10].

4. CONCLUSION

With each passing day, deepfake technology continues to advance, increasingly blurring the boundary between reality and fiction. Technology itself is neutral, and "a moral judgment can only be made about the person who uses technology in a particular way - not about the technology itself" [11, p. 25]. Nevertheless, many individuals misuse deepfake technology to create fabricated content. Owing to social media and various online platforms, photographs and videos of countless individuals are readily available without significant restrictions. Today, deepfake pornography threatens not only famous actresses and singers but also virtually any woman. As a result of the rapid development of deepfake technology, the risks of cyberbullying and various forms of fraud cannot be excluded. Artificial intelligence has advanced so rapidly within a short period of time that the human eye can no longer reliably determine whether content is authentic or fabricated. Videos of politicians appearing to say things that never actually occurred in reality frequently circulate on social media ahead of election campaigns. Such content has the potential to provoke a wide range of emotions among the public. Consequently, relations between states may be seriously damaged. The use of deepfake technology during elections has been observed in numerous countries, and the number of such cases continues to grow each year. Although deepfakes may not always have a direct impact on election outcomes, they raise serious concerns regarding public trust. Today, fake news can be created and disseminated much more easily than during the era of traditional media. Deepfake technology has also begun to be used as an auxiliary instrument of warfare. In this context, it can be employed to weaken troop morale and contribute to incorrect political assessments and decisions. The consequences may be severe, including the loss of military victories, human lives, and material resources. Detection methods are unable to keep pace with the parallel development of deepfake technology. Despite the existence of various detection techniques, some recordings may appear so authentic that it becomes difficult to determine whether and how their falsity can be proven. However, one fact remains certain: trust in information and communication systems is increasingly eroding. Today, individuals can hardly trust their own eyes. The consequences of losing confidence in the authenticity of media content are illustrated by the case from Central Africa discussed earlier in this paper. It is likely that those who initially conceived and developed this technology did not anticipate the harmful consequences that its misuse would produce within only a few years. Nevertheless, it is evident that not only individuals are endangered, but also the security of corporations and states. Therefore, public education and the adoption of appropriate legal regulations governing the use and distribution of artificially generated content are of paramount importance. In this regard, the European Union currently appears to be the most advanced actor and may serve as an example of good practice for other states seeking to address the challenges posed by deepfake technology.

References

[1] VELJKOVIĆ, S., ĆURČIĆ, M., GAVRILOVIĆ, I.: *Dark sides of deepfake technology,* Military Technical Courier/Vojnotehnički glasnik, 72(3) (2024) 1441-1463.

[2] PATEL, Y., TANWAR, S., GUPTA, R., BHATTACHARYA, P., DAVISON, I. E., NYAMEKO, R., ALUVALA, S., VIMAL, V.: *Deepfake generation and detection: Case study and challenges,* IEEE Access, 11 (2023) 143296-143323.

[3] NASIRI, S., HASHEMZADEH, A.: *The evolution of disinformation from fake news propaganda to AI-driven narratives as deepfake, Journal of Cyberspace Studies*, 9(1) (2025) 229-250.

[4] WESTERLUND, M.: *The emergence of deepfake technology: A review,* Technology innovation management review, 9(11) (2019) 39-52.

[5] YU, P., XIA, Z., FEI, J., LU, Y.: *A survey on deepfake video detection,* IET Biometrics, 10(6) (2021) 607-624.

[6] CINAR, B.: *Deepfakes in Cyber Warfare: Threats, Detection, Techniques and Countermeasures,* Asian Journal of Research in Computer Science, 16(4) (2023) 178-193.

[7] SOPHIA, L. I.: *The social harms of ai-generated fake news: Addressing deepfake and ai political manipulation,* Digital Society & Virtual Governance, 1(1) (2025) 72-88.

[8] ŁABUZ, M., NEHRING, C.: *On the way to deep fake democracy? Deep fakes in election campaigns in 2023*, European Political Science, 23(4) (2024) 454-473.

[9] KOOPMAN, M., RODRIGUEZ, A. M., GERADTS, Z.: Detection of deepfake video manipulation, in: *Proceedings of the 20th Irish Machine Vision and Image Processing Conference (IMVIP),* Irish Pattern Recognition & Classification Society, Belfast,

(2018) 133–136.

[10] TWOMEY, J., CHING, D., AYLETT, M. P., QUAYLE, M., LINEHAN, C., MURPHY, G.: *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine*, PLoS ONE 18(10), (2023) 1-22.

[11] GROTHAUS, M: *Trust No One: Inside the World of Deepfakes,* Hodder Studio, Great Britain, 2021.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A MULTI-OBJECTIVE APPROACH TO OPTIMIZING VIRTUALIZED MILITARY DATA CENTER INFRASTRUCTURE WITH GENETIC ALGORITHMS

NIKOLA ZOGOVIĆ
Mihajlo Pupin Institute, University of Belgrade, Belgrade, Republic of Serbia
nikola.zogovic@pupin.rs, ORCID: 0000-0002-9052-6422

SAVA STANIŠIĆ
Faculty of Technical Sciences, University of Kragujevac, Čačak, Republic of Serbia
sava.stanisic@vs.rs, ORCID: 0009-0002-3118-0537

BORISLAV ĐORĐEVIĆ
Mihajlo Pupin Institute, University of Belgrade, Belgrade, Republic of Serbia
borislav.djordjevic@pupin.rs, ORCID: 0000-0002-6145-4490

Abstract: *Selecting cost-efficient, highly available, and reliable virtualized infrastructure is a critical challenge for military data centers. This paper extends a prior two-objective optimization framework by introducing reliability as a third objective alongside cost minimization and availability maximization. The framework uses the Non-dominated Sorting Genetic Algorithm II (NSGA-II) to model trade-offs among three conflicting objectives within a virtualized military data center, where services are deployed as virtual machines across physical nodes with heterogeneous failure characteristics. Five VM types with distinct cost, availability, and failure rate profiles are defined to model realistic military infrastructure. Experimental results with the pymoo library demonstrate that the three-objective approach yields 119 Pareto-optimal solutions spanning \$0.18–\$4.94/h in cost, 98.0–100.0000% in availability, and 23.7–99.9996% in reliability over a 720-hour horizon. Critically, at matched cost points, the three-objective approach achieves up to 20.5 percentage points higher reliability than the two-objective baseline by diversifying VM type selection rather than stacking identical low-cost instances.*

Keywords: *multi-objective optimization, NSGA-II, Pareto optimality, military data center, virtualization, availability, reliability*

1. INTRODUCTION

Modern military operations depend on robust digital infrastructure to support command and control (C2), intelligence, surveillance, and reconnaissance (ISR), logistics management, and secure communications [1, 2]. These services are increasingly deployed on virtualized platforms within military data centers, both in garrison facilities and in field-deployable configurations. Unlike commercial cloud environments where downtime results in financial loss, failures in military IT infrastructure can compromise mission execution, situational awareness, and force protection [3].

Military data centers operate under unique constraints. They are often isolated from public networks, must function in austere or contested environments, and face rigid procurement budgets. The selection of virtual machine (VM) types and their allocation across physical servers is a critical planning decision that directly affects operational readiness. From the requirement to ensure uninterrupted support for command, control, and intelligence functions, availability and reliability arise as fundamental system objectives. Availability, the probability that a system is operational, determines acceptable outage duration [4]. Reliability, the probability of continuous failure-free operation over time T, is equally critical, as even brief C2 or ISR interruptions during operations have tactical consequences [5].

Traditional single-objective optimization approaches fall short in addressing this multi-faceted problem. The authors proposed a two-objective NSGA-II framework in [6] to balance cost and availability. However, it did not explicitly optimize reliability, instead assuming that higher availability implies higher reliability, an assumption that fails when VM types have heterogeneous failure characteristics.

This paper extends [6] by incorporating reliability as a third objective and adapting the optimization framework to military data centers. The results show that optimizing only cost and availability systematically favors homogeneous VM deployments, typically using multiple low-cost instances to maximize availability while reducing reliability because of their higher failure probability. Similar to a military force relying on a single weapon system, such configurations are efficient but vulnerable to type-specific weaknesses. Adding reliability

Corresponding author: N. Zogović;
E-mail: nikola.zogovic@pupin.rs
DOI: 10.5937/OTEH26079Z

as a third objective diversifies VM portfolios by failure-rate profile, providing a computational analogue of defense in depth. This optimization increases modeled reliability by up to 20.5 percentage points at the same cost.

This paper makes three contributions: (1) it introduces reliability as a third optimization objective for virtualized military data centers; (2) it defines a five-tier VM model capturing the heterogeneous failure characteristics of military-grade hardware; and (3) it demonstrates that VM type diversification, rather than redundant replication of identical instances, improves reliability. Unlike prior multi-objective cloud studies that optimize cost and availability or task scheduling without reliability [6–10], the proposed formulation explicitly captures the cost–availability–reliability trade-off, providing actionable guidance for infrastructure planners.

2. RELATED WORK

Multi-objective optimization (MOO), particularly using NSGA-II, has been widely applied to cloud infrastructure optimization. Tan et al. [7] applied NSGA-II to microservice allocation in containerized clouds, while Verma et al. [8] identified it as a state-of-the-art method for Pareto-based combinatorial optimization.

In military computing, virtualization is a key enabler of deployable C2 systems and tactical data centers. Fault-tolerant VM placement for redundant server clusters was studied in [3], [9] presents NSGA-II application to cloud task scheduling, while [10] presents heterogeneous IoT–fog–cloud environments. More recent work [11,12] has extended multi-objective, QoS-aware fault-tolerant VM placement to data-center-scale systems.

In previous work [6], the authors proposed a two-objective NSGA-II framework balancing cost and availability, and Stanišić [13] investigated Docker file-system performance for containerized workloads. However, existing studies do not explicitly optimize reliability, often consider only a limited range of VM types, and largely overlook VM type diversification as a mechanism for improving system resilience. This paper addresses these gaps by introducing reliability as a third optimization objective and modeling heterogeneous VM failure characteristics.

3. METHODOLOGY

The optimization framework uses NSGA-II [14] to simultaneously minimize cost and maximize availability and reliability for a virtualized military data center.

3.1. Problem Formulation and Military VM Tier Model

The decision vector $x = [n_1, n_2, n_3, n_4, n_5]$ represents the number of VM instances of each type to deploy, where $n_i \in \{0, 1, \ldots, 10\}$. The uniform upper bound of 10 is a scenario-specific capacity proxy based on data center resource constraints (compute, memory, space, and power). For operational use, it must be replaced by tier-specific limits derived from the actual host inventory. Five VM types are defined to model the range of hardware typically available in military data centers, from commercial off-the-shelf (COTS) servers used for administrative workloads to hardened, fault-tolerant platforms designed for mission-critical C2 systems. The specifications are shown in Table 1.

Table 1. Military Data Center VM Tier Specifications

VM Tier	Cost ($/h)	Avail.	λ (/h)	Typical Use
T1: COTS Basic	$0.18	0.980	0.002	Admin, logs
T2: COTS Enhanced	$0.37	0.990	0.001	Logistics
T3: Hardened	$0.52	0.995	0.0005	SIGINT proc.
T4: Mission-Crit.	$0.78	0.999	0.0002	ISR feeds
T5: Fault-Tolerant	$1.15	0.9995	0.0001	C2 systems

The tier model reflects real-world military procurement trade-offs. Tier 1 (COTS Basic) represents commodity servers suitable for non-critical workloads such as administrative systems and log aggregation. Tiers 3–5 represent progressively hardened platforms with redundant power supplies, ECC memory, and hot-swappable components, designed for signal intelligence processing, ISR data feeds, and command and control systems respectively. The failure rate λ decreases by an order of magnitude from T1 to T5, reflecting the engineering investment in higher-tier hardware.

The failure rate parameters are derived from representative mean time between failures (MTBF) values: T1 corresponds to MTBF ≈ 500h (commodity COTS hardware), T2 to MTBF ≈ 1000h (enterprise-grade with ECC RAM), T3 to MTBF ≈ 2000h (ruggedized platforms, e.g., MIL-STD-810 compliant), T4 to MTBF ≈ 5000h (mission-critical with redundant PSU and cooling), and T5 to MTBF ≈ 10000h (fault-tolerant systems with hot-standby components). Availability values are derived from failure rates assuming a mean time to repair (MTTR) of 1 h (T1–T2) and 0.25 h (T3–T5), reflecting the faster repair of modular, hot-swappable hardware in higher tiers.

3.2. Objective Functions

Three objectives are defined. All objectives are formulated as minimization for NSGA-II compatibility.

Objective 1: Cost Minimization. The total infrastructure cost is the sum of per-instance costs:

$$C(x) = \sum_i n_i \cdot c_i \tag{1}$$

where c_i is the hourly cost of VM type i, given in column 2 of Table 1. The optimization minimizes $f_1 = C(x)$.

Objective 2: Availability Maximization. Service availability is modeled using the classical parallel-system

availability model, assuming independent VM failures. The probability that at least one VM remains operational is

$$A(x) = 1 - \prod_i (1 - a_i)^{n_i} \quad (2)$$

where a_i is the per-instance availability of VM type i (Table 1). The optimization minimizes $f_2 = 1 - A(x)$.

Objective 3: Reliability Maximization. System reliability over the time horizon T is modeled using the classical exponential lifetime model and parallel-system reliability assumption, yielding the probability that at least one VM remains operational:

$$R(x) = 1 - \prod_i (1 - e^{-\lambda_i \cdot T})^{n_i} \quad (3)$$

where λ_i is the failure rate of VM type i (Table 1), and $T = 720\text{h}$ (30 days). The time horizon is mission-dependent and reflects the operation or deployment duration (Table 3). The optimization minimizes $f_3 = 1 - R(x)$. Unlike availability, reliability depends on the failure rate λ_i rather than the availability a_i; consequently, a VM type may exhibit high availability through rapid recovery yet low reliability due to frequent failures, or vice versa.

Equations (2) and (3) assume independent VM failures and omit common-cause failures from shared infrastructure. Thus, the calculated availability and reliability may be optimistic when instances share a physical fault domain. Incorporating common-cause failures and fault-domain placement is left for future work.

3.3. Constraints and Problem Statement

The optimization is subject to a budgetary constraint $C(x) \leq B$, where B is the predefined operational budget. Two scenarios model distinct military contexts: $B = \$5/h$ represents a field-deployable tactical data center with limited power and hardware, while $B = \$15/h$ represents a garrison facility with greater infrastructure capacity. The per-type instance count is bounded as described in Section 3.1; no explicit floor is imposed on availability or reliability, so the reported Pareto fronts span the full feasible range of these objectives.

Combining the three objectives with the budget and instance-count constraints, the complete optimization problem is stated as:

$$x* = argmin_x \left(f_1(x), f_2(x), f_3(x)\right), \; s.t. \; C(x) \leq B, \; n_i \in \{0, \ldots, 10\} \quad (4)$$

Each component n_i takes one of 11 integer values, so the decision space comprises $11^5 = 161{,}051$ candidate configurations, over which NSGA-II searches for the Pareto-optimal set.

3.4. Algorithm Configuration

NSGA-II is configured with a population size of 120, 300 generations, simulated binary crossover (SBX; probability 0.9, η = 15), polynomial mutation (probability 0.2, η = 20), and integer-rounding repair. For the algorithm comparison in Section 4.6, NSGA-III is configured with the same population size, generation count, and genetic operators, using Das–Dennis reference directions to guide selection. The implementation uses the pymoo library (v0.6.1). Hypervolume convergence is tracked to verify algorithm convergence.

4. EXPERIMENTAL RESULTS

4.1. Two-Objective Baseline

The two-objective optimization (cost vs. availability), the problem originally defined in our prior work [6], under the \$5/h budget produces 10 Pareto-optimal solutions. Fig. 1 shows the Pareto front.

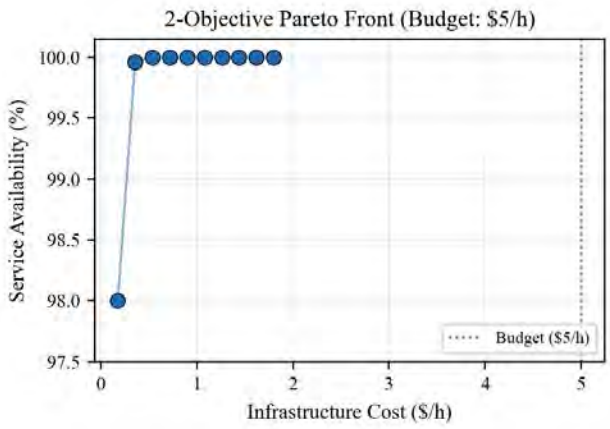


Figure 1. 2-Objective Pareto Front (Budget: \$5/h)

A critical observation is that all solutions consist exclusively of GP-Low instances (the cheapest VM type). The algorithm maximizes availability by stacking identical low-cost VMs, ranging from 1 instance (\$0.18/h, 98.0%) to 10 instances (\$1.80/h, 100.0%). This monoculture strategy is optimal for availability but ignores reliability implications.

4.2. Three-Objective Optimization

The three-objective optimization, defined by the objectives in Eqs. (1)–(3) and the constraints of Section 3.3, under the $\$5/h$ budget yields 119 Pareto-optimal solutions spanning cost $\$0.18 - \$4.94/\text{h}$, availability $98.0 - 100.0000\%$, and reliability $23.7 - 99.9996\%$ (T = 720h). Fig. 2 shows the 3D Pareto front.

The Pareto surface reveals the three-way trade-off. Low-cost solutions (bottom-left) sacrifice both availability and reliability. As cost increases, both metrics improve, but the rate of improvement differs, with reliability showing steeper gains at moderate cost levels due to the exponential nature of the failure rate model.

Under the relaxed budget of $\$15/\text{h}$, the Pareto front extends to 119 solutions with reliability reaching 100.00% (Fig. 3).

4.3. Key Finding: Reliability Improvement over the Two-Objective Baseline through VM Diversification

The central finding of this study is revealed by comparing 2-objective and 3-objective solutions at matched cost points. Fig. 4 presents this comparison and Fig. 5 shows the cost-reliability projection for both budget scenarios.

The mechanism is clear: the 2-objective approach, optimizing only cost and availability, consistently selects homogeneous configurations of the cheapest VM tier (T1: COTS Basic, $\lambda = 0.002/\mathrm{h}$). While stacking many T1 instances achieves near-perfect availability through redundancy, the high per-instance failure rate degrades reliability over the 720h (30-day) evaluation horizon, a duration representative of a typical military deployment rotation or exercise cycle.

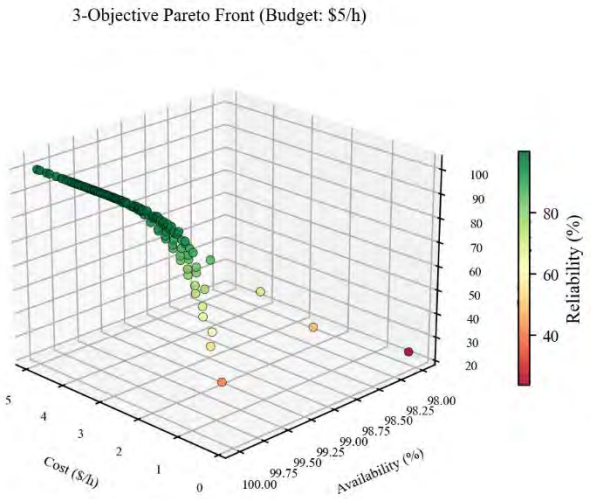


Figure 2. 3-Objective Pareto Front (Budget: $5/h)

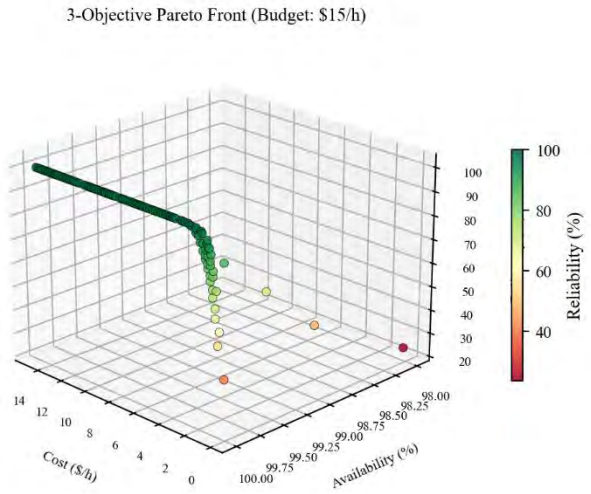


Figure 3. 3-Objective Pareto Front (Budget: $15/h)

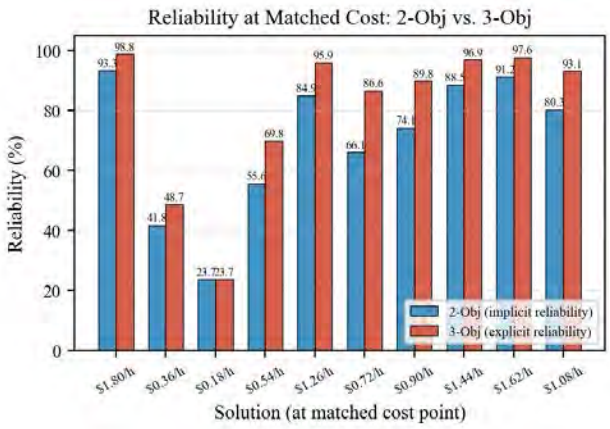


Figure 4. Reliability Comparison at Matched Cost Points

The 3-objective approach discovers that mixing VM tiers, particularly by including T3: Hardened ($\lambda = 0.0005/\mathrm{h}$) and T4: Mission-Critical ($\lambda = 0.0002/\mathrm{h}$) instances, dramatically improves reliability. For example, at $\$0.72/\mathrm{h}$, the 2-objective solution deploys 4×T1 instances and achieves 66.1% reliability, while the 3-objective solution deploys 1×T4 instance and achieves 86.6%, a relative gain of 20.5 percentage points over the two-objective baseline at the same cost. In military terms, this is the difference between a C2 system that is expected to fail multiple times during a 30-day deployment and one that has an 87% probability of continuous operation.

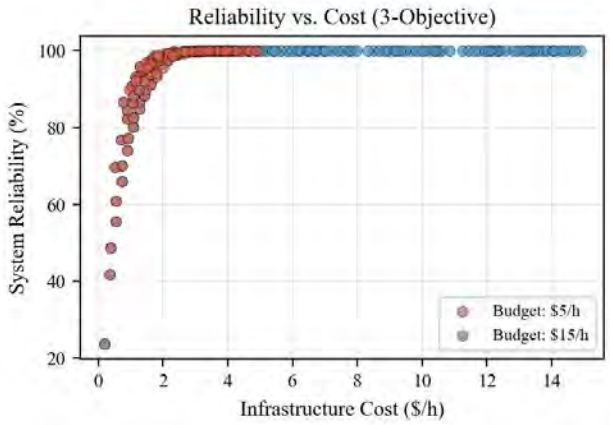


Figure 5. Reliability vs. Cost Projection

Table 2. Reliability Improvement: 2-Obj vs. 3-Obj at Matched Cost

Cost ($/h)	2-Obj Rel.(%)	3-Obj Rel.(%)	Δ Rel.(pp)
$0.54	55.6	69.8	+14.2
$0.72	66.1	86.6	+20.5
$1.08	80.3	93.1	+12.9
$1.26	84.9	95.9	+11.0
$1.44	88.5	96.9	+8.4

This result is a computational manifestation of the defense-in-depth principle: heterogeneous systems are more resilient than homogeneous ones because they are not susceptible to common-mode failures. A vulnerability or hardware defect affecting T1 COTS servers would

simultaneously compromise all 4 instances in the 2-objective solution, while the 3-objective solution's T4 mission-critical hardware operates on a fundamentally different failure profile.

4.4. VM Type Allocation Patterns

Fig. 6 shows how VM type allocation varies across the Pareto front.

Low-cost Pareto solutions use T1 COTS instances exclusively, mirroring the 2-objective behavior. As budget increases, the optimizer progressively introduces higher-tier hardware. At moderate cost levels ($\$0.52 - \$1.30/h$), the algorithm discovers mixed configurations combining T1 and T3/T4 instances, providing both the redundancy benefits of multiple instances and the low failure rates of hardened hardware. At higher budgets, T4 Mission-Critical and T3 Hardened instances dominate, reflecting the military planning principle that mission-critical services should run on mission-critical hardware when resources permit.

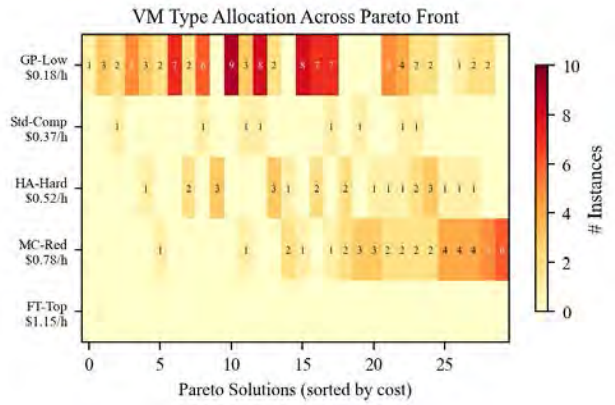


Figure 6. VM Type Allocation Across Pareto Front

4.5. Sensitivity Analysis: Effect of Time Horizon

A key question for military planners is how the operational time horizon affects the value of three-objective optimization. Figure 7 shows the reliability gap between the two- and three-objective formulations across four operationally relevant time horizons.

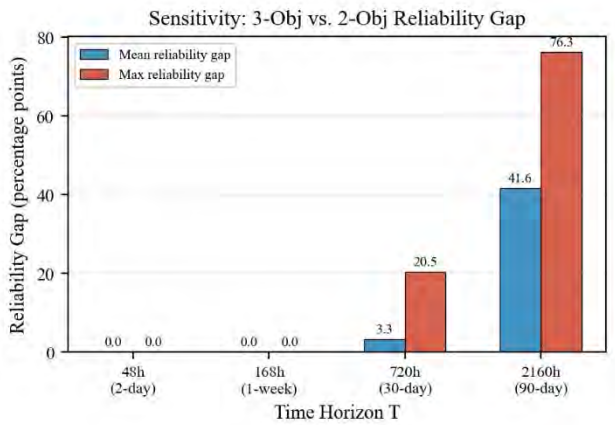


Figure 7. Reliability Gap vs. Time Horizon

The results reveal a clear pattern. For short operations ($T \leq 168$h), all VM types remain sufficiently reliable, and the two-objective formulation performs as well as the three-objective approach. As the time horizon increases to 720h (30-day deployment), a reliability gap emerges, reaching up to 20.5 percentage points. At $T = 2160h$ (90-day rotation), the gap becomes substantial, with an average of 41.6 and a maximum of 76.3 percentage points. These results indicate that while explicit reliability optimization provides little benefit for short missions, it becomes essential for sustained deployments exceeding one week.

4.6. NSGA-II vs. NSGA-III comparison

Fig. 8 compares the convergence of NSGA-II and NSGA-III on the three-objective optimization problem. NSGA-II produces 119 Pareto-optimal solutions with a final hypervolume of 14.45, whereas NSGA-III produces 12 solutions with a hypervolume of 14.26. NSGA-II performs better because NSGA-III is designed for many-objective optimization (four or more objectives), where reference directions are needed to maintain diversity. In contrast, the present three-objective problem has a small, bounded integer search space, allowing NSGA-II's crowding-distance mechanism to preserve diversity more effectively. These results justify the use of NSGA-II for military data center optimization with integer VM allocation variables.

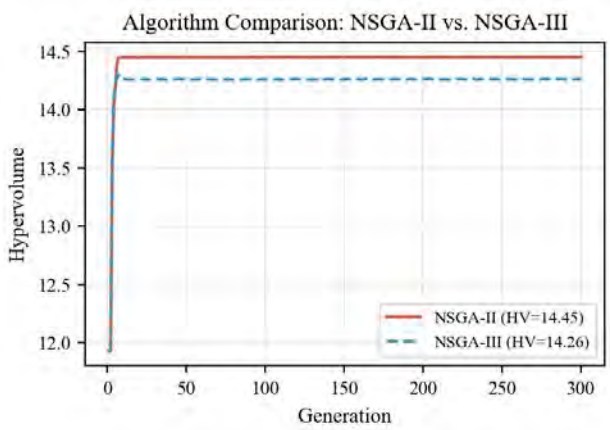


Figure 8. NSGA-II vs. NSGA-III Convergence

Hypervolume analysis further shows that both algorithms reach 99% of their final hypervolume within the first few tens of generations. The three-objective formulation converges as quickly as the two-objective formulation, indicating that adding reliability as a third objective incurs negligible computational overhead.

4.7. Full Objective Space and Objective-Level Constraints

Sections 4.1–4.2 present only Pareto-optimal solutions. To characterize the search space more completely, all $11^5 = 161{,}051$ VM configurations were exhaustively evaluated using Eqs. (1)–(3). Fig. 9 shows the resulting objective space projected onto the cost–reliability plane for the tactical budget (B = \$5/h), under which 2,503 configurations satisfy the budget constraint. Reliability

ranges from 23.7% (a single T1 COTS instance) to 99.9996% (highly redundant mixed configurations), demonstrating that the budget constraint alone permits configurations with widely different mission suitability.

The formulation in Section 3.3 imposes no minimum reliability, so all budget-feasible configurations are included in the Pareto analysis. In practice, however, missions often require a minimum reliability. Applying an objective-level constraint, $R(x) \geq R_{min}$, removes the lower portion of the objective space. For a three-nines requirement ($R_{min} = 99.9\%$ over the 720h horizon, dashed line in Fig. 9), 625 of the 2,503 budget-feasible configurations (25.0%) are eliminated, leaving 1,878 feasible configurations. A four-nines requirement ($R_{min} = 99.99\%$) excludes 1,582 configurations (63.2%). Most eliminated configurations are the low-cost, low-redundancy monocultures identified in Section 4.3.

These results show that realistic mission-reliability requirements eliminate many of the cheapest solutions, leaving diversified, higher-reliability VM portfolios as the feasible design choices. When mission reliability is specified a priori, the threshold can be incorporated directly as an additional optimization constraint.

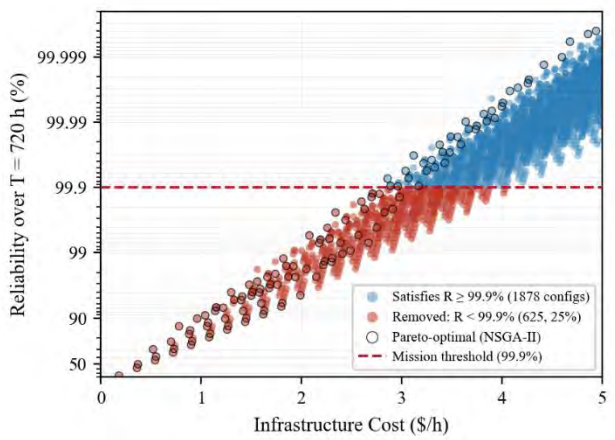


Figure 9. Full objective space for B = $5/h

5. CONCLUSION

This study extends the two-objective cloud optimization framework in [6] by introducing reliability as a third objective for virtualized military data centers. The results show that optimizing only cost and availability systematically favors homogeneous COTS-tier VM deployments, which achieve high availability through redundancy but poor long-term reliability. Incorporating reliability fundamentally changes the optimal solutions, with diversified VM portfolios combining COTS, hardened, and mission-critical hardware improving reliability by up to 20.5 percentage points at the same cost.

The findings provide three practical recommendations. First, budget-constrained tactical data centers should deploy mixed T1–T4 VM configurations rather than replicate identical low-cost instances. Second, higher-budget garrison facilities should reserve T3–T5 hardware for mission-critical workloads while using T1–T2 for non-critical services. Third, the resulting Pareto front provides a decision-support tool for balancing cost, availability, and reliability according to mission priorities.

Future work will extend the framework by incorporating adversarial failure models, workload-specific reliability targets, anti-affinity constraints for physical fault-domain separation, complete analysis of the objective space, and AI-based adaptive reconfiguration [15] for dynamic optimization under changing mission conditions.

References

[1] MALHOTRA, R., BANSAL, A., KESSENTINI, M.: A systematic literature review on maintenance of software containers, ACM Computing Surveys, 56(8), pp. 1–38, 2024.

[2] SMITH, W. et al.: Cloud computing in tactical environments, Proc. MILCOM 2017 – IEEE Military Communications Conference, pp. 882–887, 2017.

[3] MACHIDA, F., KAWATO, M., MAENO, Y.: Redundant virtual machine placement for fault-tolerant consolidated server clusters, IEEE/IFIP NOMS, pp. 32–39, 2010.

[4] NABI, M., TOEROE, M., KHENDEK, F.: Availability in the cloud: State of the art, J. Network and Computer Applications, 60, pp. 54–67, 2016.

[5] TRIVEDI, K.S., KIM, D.S., GHOSH, R.: Resilience in computer systems and networks, Proc. IEEE/ACM ICCAD, pp. 74–80, 2009.

[6] STANIŠIĆ, S., ZOGOVIĆ, N.: A Multi-Objective Approach to Optimizing Cloud Infrastructure with Genetic Algorithms, Proc. 12th IcETRAN, pp. 1–4, 2025.

[7] TAN, B., MA, H., MEI, Y.: A NSGA-II-based Approach for Multi-objective Micro-service Allocation in Container-based Clouds, Proc. IEEE/ACM CCGRID, pp. 282–289, 2020.

[8] VERMA, S., PANT, M., SNASEL, V.: A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems, IEEE Access, 9, pp. 56745–56777, 2021.

[9] ARORA, M., KUMAR, V., DAVE, M.: Task Scheduling in Cloud Infrastructure using Optimization Technique Genetic Algorithm, Proc. 4th WorldS4, pp. 788–793, 2020.

[10] MOKNI, I., YASSA, S.: A multi-objective approach for optimizing IoT applications offloading in fog-cloud environments with NSGA-II, The Journal of Supercomputing, 80, pp. 27034–27072, 2024.

[11] ZHANG, W., CHEN, X., JIANG, J.: A multi-objective optimization method of initial virtual machine fault-tolerant placement for star topological data centers of cloud systems, Tsinghua Science and Technology, 26(1), pp. 95–111, 2021.

[12] BALAJI, K. et al.: Virtual Machine Fault Tolerance Placement in Data Centers using Multi-Objective Optimization, Proc. 8th CSITSS, pp. 1–6, 2024.

[13] STANIŠIĆ, et al.: Performance Optimization of File Systems for Docker Containers, Proc. Int. Scientific Conf. Sinteza 2025, pp. 117–127, 2025.

[14] SHI, L., DUAN, S.: Design and Implementation of Improved Multi-objective Genetic Algorithm Based on Uniform Distribution, DIONE 2023, LNICST, vol. 515, Springer, 2025.

[15] STANIŠIĆ, S., LUKOVIĆ, V., BELOTIĆ, B.: Automation of Monitoring and Optimization of Docker Containers Using Artificial Intelligence, Proc. 24th Int. Symp. INFOTEH-JAHORINA, pp. 1–4, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A MULTI-CRITERIA APPROACH FOR AUTOMATION OF REAL-TIME WEAPON ASSIGNMENT PROCESS

DARKO PIJEVIĆ
Military Technical Institute, Belgrade, Serbia, darko.pijevic@gmail.com

KSENIJA BOKIĆ
Institute Mihajlo Pupin, Belgrade, Serbia, ksenija.bokic@pupin.rs

INA MASNIKOSA
Institute Mihajlo Pupin, Belgrade, Serbia, ina.masnikosa@pupin.rs

TANJA LACMANOVIĆ
Institute Mihajlo Pupin, Belgrade, Serbia, tanja.lacmanovic@pupin.rs

Abstract: *This paper addresses the Threat Evaluation and Weapon Assignment (TEWA) problem in Ground-Based Air Defence (GBAD) systems. GBAD systems incorporate sensors, effectors, command and control (C2), and telecommunication systems to provide efficient short-to-medium-range air defence. At the functional level, GBAD systems are designed to detect, track, and engage aerial targets. The decision-making problem in GBAD systems is characterised by short response time and human factors. Combined with limited engagement time, automation of the decision-making process becomes imperative. In modern-day asymmetric warfare conditions, GBAD systems must provide a high level of efficiency and automation in the decision-making process. This paper proposes an approach based on the Find, Fix, Track, Target, Engage and Assess (F2T2EA) executive procedure. Key points in the F2T2EA executive procedure are TEWA, performed in real time, with efficient and precise estimation of aerial objects' capability and intent to engage Defended Assets (DAs). The algorithm is based on layered multi-criteria decision-making analyses (MCDMA), implemented and tested in a simulated environment, showing overall effectiveness. Compared to the Baseline algorithm, which simulates fast human response, the proposed algorithm shows a ~46% decrease in total scenario execution time and a ~28% decrease in total threat exposure of DA over time, demonstrating a significant increase in DA survivability.*

Keywords: *threat evaluation, weapon assignment, GBAD systems, MCDMA*

1. INTRODUCTION

Ground-Based Air Defence (GBAD) systems incorporate sensors, effectors, command and control (C2), and telecommunication systems to provide efficient short- to medium-range air defence. Unlike long-range air defence systems (LRAD), which primarily deal with medium- to high-altitude and high-speed targets (such as fighters, bombers, medium-range ballistic missiles, etc.), short-range air defence systems (SHORAD) deal with low-altitude, low radar cross-section and high-speed, or low/slow/small threats (such as loitering munitions, drone swarms, and sea-skimming cruise missiles). Thus, LRAD systems' reaction time is measured in minutes, and SHORAD systems' reaction times drop from minutes to seconds.

To enable efficient resource management of available effectors and ammunition vs. designated targets, the Find→ Fix→ Track→ Target→ Engage→ Assess (F2T2EA) executive procedure is applied. This paper focuses on the Target→ Engage→ Assess steps of F2T2EA, while the Find→ Fix→ Track steps are simulated and hypothetical. Another focus of this paper is the automation of the decision-making process in order to provide maximum efficiency and survivability of the Defended Asset (DA) while reducing the need for human-in-the-loop in the mentioned process [1]. In modern-day asymmetric warfare conditions, SHORAD systems must provide a high level of efficiency and automation in the decision-making process. Threat Evaluation and Weapon Assignment (TEWA) problem is a critical decision support process in air defence and military C2 platforms. It typically consists of two mutually independent but interrelated sub-problems:

1. TE: Quantifying and prioritising incoming aerial or missile threats based on factors like intent, velocity, trajectory, distance, and DA priority (criticality).
2. Weapon Assignment (WA): Dynamically allocating optimal weapon systems to neutralise previously prioritised threats in real-time, under defined operational constraints.

The proposed TEWA system uses an algorithm based on Multi-Criteria Decision-Making Analyses (MCDMA) to produce efficient engagement planning and target distribution [2].
Simulation is performed following the Stage-Based Weapon Target Assignment methodology [3]. The efficiency assessment criterion is based on a cross-reference evaluation framework by defining appropriate

Corresponding author: D. Pijević;
E-mail: darko.pijevic@gmail.com
DOI: 10.5937/OTEH26080P

Key Performance Indicators (KPI) [4]. Efficient TEWA reflects military proficiency and requires prompt decision-making while managing the available resources [2].

2. LITERATURE REVIEW AND METHODOLOGICAL FRAMEWORK

2.1. F2T2EA execution cycle (kill-chain)

F2T2EA execution cycle enables the entire end-to-end operational cycle to provide efficient resource management from target acquisition to engagement results assessment. The F2T2EA kill chain framework is the standard process used for dynamic targeting and time-sensitive threat response in modern joint military operations [5]. In SHORAD and GBAD operational environments, the F2T2EA kill chain undergoes fundamental structural changes. The F2T2EA execution cycle applies not just to kinetic means, but also to military capabilities that achieve lethal or nonlethal effects through non-kinetic means [6]. Due to limited resources available (number of available rockets/projectiles, number of available channels for engagement per Fire Unit (FU), jammer saturation, FU readiness level and engagement time, FU recovery time after engagement, etc.), the F2T2EA execution cycle is considered to operate in a strictly dynamic environment. For example, the US Air Force follows the F2T2EA kill chain for the Dynamic Targeting Process (DTP) [6]. F2T2EA DTP execution cycle is depicted in Figure 1.

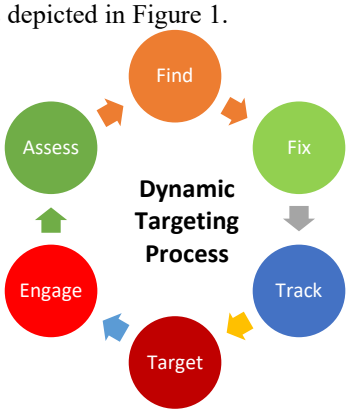


Figure 1 F2T2EA DTP Execution Cycle

Main characteristics of the F2T2EA DTP kill-chain application in SHORAD systems are the following:

- Linearity and dependency – each phase is directly dependent on results obtained from the previous phase [6];
- Time criticality – the cycle must be designed to meet requirements for dynamic target kinematics combined with short time for engagement [7];
- Iterative – direct output from the *assess* phase is taken into consideration for future FUs resource management [8].

Technically, if the *track* phase of F2T2EA fails, no target designation and engagement is possible. Also, if results of the *assess* phase are negative (e.g., target is missed), FU resources remain occupied for further engagement on the designated target. Thus, F2T2EA DTP practically provides a correct and efficient way for resource management over available FUs.

2.2. Threat Evaluation (TE)

In military operations, TE measures how dangerous an entity (aircraft, missile, drone, or other airborne vessel) is to DA at a given point in time. It converts raw tracking and sensor data into a continuous numerical threat score or a ranked priority list, updated in real time (often on every tracking data refresh). With respect to the object of TE, two base methods are known: DA-based TE and target-based TE. With respect to methodology applied, three classes of TE can be found in the literature [8],[9],[10]:

1. Analytical Methods: an approach using strict mathematical equation programming for TE execution – achieving global optimality and predictive system behaviour;
2. Heuristic & Metaheuristic Methods: an approach that skips guaranteed global optimality in exchange for calculating "fast and good enough" threat scores;
3. AI & Machine Learning Methods – an approach that focuses on handling uncertainty, incomplete radar tracking data, target deception, and non-linear intent inference.

2.3. Weapon Assignment (WA)

In SHORAD systems, Weapon Assignment (WA) – often designated as the Weapon-Target Assignment (WTA) problem – is the optimisation engine of the TEWA cycle. Its objective is to optimally allocate available defensive effectors (missiles, anti-aircraft guns, jammers, directed energy weapons) to incoming aerial threats and secure minimal DA damage, while strictly satisfying operational constraints (ammunition inventory, launcher slewing, fire channels). According to [11], WA refers to the reactive assignment of FUs to engage identified threats in a defined operational timeline, consistent with the related own force mission objectives and compliant with the Rules of Engagement (ROE) [8],[12]. The WA problem may be considered from a number of different perspectives [8],[12]. When investigating which FUs to assign, one of the following two perspectives is usually adopted:

- The single platform perspective – referring to a single platform protecting itself from threats, and the main problem is selecting the most suitable weapon system,
- The force coordination perspective – referring to a C2 platform providing TE against a predefined set of DAs, where assessment relates to the identification of the most suitable FUs to engage or counter a threat.

Furthermore, when investigating the approach by which FUs are assigned, two perspectives are prevalent [5],[8],[12]:

- The threat-by-threat perspective, which refers to the assignment of FUs sequentially, in such a way that the best FU is essentially assigned to each threat in turn (from the highest priority to the lowest priority),
- The multi-threat perspective, which refers to the assignment of FUs to the current set of threats concurrently, so that the assignment is best in some overall sense.

In terms of methodology used, major WTA method families in GBAD and SHORAD systems follow [8],[13],[14],[15]:

- Exact / Integer solvers - solving the FU allocation optimisation under fixed constraints,
- Heuristics and Greedy methods - rule-based heuristics immediately assign available FUs to the highest-ranked targets in real time, guaranteeing solution feasibility without iterative loops,
- Bi-level Pareto Evolutionary - minimises DA damage while minimising missile inventory depletion, producing a set of non-dominated Pareto-optimal trade-off solutions,
- Game-Theoretic and Multi-Agent - evaluates weapon assignment in electronic warfare environments or dynamic salvo scenarios, treating the engagement as a multi-stage game with uncertain target tracks,
- Machine learning – primarily supervised and reinforcement learning.

3. THE METHOD

3.1. Methodology selection

Due to the problem statement, the decision-making process in SHORAD and GBAD systems must reside on a methodology that is a compromise between computational complexity, the desired optimal solution, and the possibility for real-time application across multiple DAs and numerous potential threats, simultaneously. Therefore, the TE part, as an entry point, must correctly analyse potential threat behaviour with respect to a predefined set of DAs, and the WA part must provide efficient optimisation of target designation across a set of available FUs.

With respect to the short reaction time of SHORAD systems and reduced scope of targets for engagement, this paper focuses on DA-based TE, with an analytical methodology for kinematic threat analyses of potential targets. A standard DA-based analytical kinematic TE model evaluates three main dimensions [8],[11]:

1. Intent (Q: What can the target do?) – based on target type and kinematic capability for engagement;
2. Trajectory (Q: Where is it going?) – based on estimation of Closest Point of Approach (CPA) and Time to CPA;
3. DA Priority Value / Criticality (Q: What is at risk?) – not all DAs are equal, and it is mandatory to define a priority parameter for each DA.

To optimise available resources across critical/priority targets obtained by TE, while maintaining moderate computational complexity, choosing between heuristics and MCDMA is considered justified. This paper addresses applications of the heuristic “Greedy” model and MCDMA model for optimising the decision-making process. To perform a fair efficiency analysis of the obtained results, a suitable set of KPIs must be defined. Given the SHORAD and GBAD systems' operational objectives, ensuring the survivability of DA is an absolute priority. Therefore, a set of KPIs for cross-referencing and analysis can be defined as follows:

- Integral of Total Threat Exposure (ITTE) of DA over time – representing the area under the total threat score curve during Total Scenario Execution Time (TSET),
- TSET – representing the time period in which all high-intensity manoeuvring targets are neutralised/destroyed.

Using the proposed set of KPIs, maximising the survivability of DA is fairly depicted through measurable numeric indicators, which essentially represent processed outputs from the proposed TEWA system, indicating its efficiency.

3.2. TOPSIS

MCDMA remains a central theme in contemporary research due to its capacity to resolve complex challenges in various fields, so it is widely used for military purposes [2],[16]. One commonly used MCDMA method in TEWA is the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). TOPSIS ranks alternatives based on their proximity to ideal and anti-ideal solutions [17]. Although the method accommodates both quantitative and qualitative data, its reliability can be affected by the weighting of criteria and the specific choice of reference points.

The criteria chosen for TOPSIS are listed and explained in Table 1.

Table 1 TOPSIS criteria

CRITERION	TYPE	OPERATIONAL IMPLICATION
Threat score (w_1)	Benefit	Prioritises engagement of the highest-threat targets to minimise exposure.
Time to destruction of target (w_2)	Cost	Favours pairs that minimise intercept duration.
Time to FU range (w_3)	Cost	Chooses targets that enter the range of FU faster.
Number of potential targets (w_4)	Cost	Favours FU with fewer potential targets.

3.3. Baseline algorithm

The Baseline algorithm serves as a reference point and counter-example to the advanced MCDMA model. It demonstrates the decision-making process based solely on one criterion: the distance from the target to the FU. Unlike the MCDMA approach, the Baseline algorithm uses a heuristic “Greedy” strategy and simulates a quick human reaction.

4. SIMULATION

To test the proposed algorithm, a realistic GBAD scenario is created. The scenario total duration is 120 seconds.

4.1. A realistic SHORAD scenario

The scenario intends to mimic a typical military exercise with a circular defence of a high-value DA. The scenario was created based on the work in [18]. In the scenario, there are three short-range FUs arranged in a triangle formation, as shown in Figure 2. Their maximum range of 8 km overlaps, and the DA is placed at the centre.

Five threats (T1-T5 in Figure 2), which represent high-intensity manoeuvring targets, enter the defended airspace. Threats approach from north-west, south-west, and south at initial ranges of ~9.5 km, and their trajectories consist of multi-segment waypoints with linear interpolation between nodes. Speeds range from 160 to 200 m/s, and altitudes vary from 2000 to 3400 m. Target T5 executes a sharp left U-turn above the DA to simulate a so-called toss bomb flight path attack technique. Ten additional targets (T6-T15 in Figure 2) organised into directional groups with staggered launch offsets simulate lower-priority background traffic. These targets follow linear paths toward the DA region. Three fixed flight profiles are used:

- High (altitude of 5000 m, speed of 250 m/s),
- Medium (altitude of 3500 m, speed of 120 m/s), and
- Low (altitude of 1700 m, speed of 40 m/s).

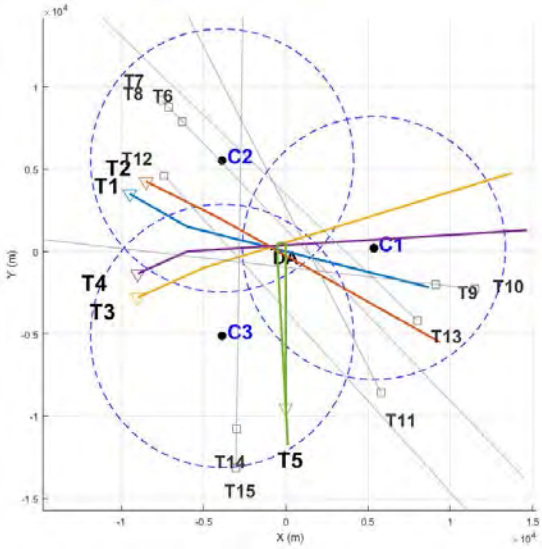


Figure 2 SHORAD/GBAD scenario

4.2. Radar data acquisition and threat score calculation

Radar data acquisition simulates a rotating air surveillance system with a fixed rotation period of 1,5 s. Additional radar parameters define the mean values and standard deviations for range, azimuth, and elevation measurements.

The threat scores quantify the immediate threat level for each detected track on a scale of 0% to 100%. The score of the represented DA is based on predefined thresholds; four threat properties are scored:

1. Direction – Smaller angular deviation yields higher scores (0°–3° with score 10, descending to score 2 for 70°–180°).
2. Speed – Prioritises fast interceptors (150–200 m/s with score 10), with reduced scores for extremely slow or hypersonic targets.
3. Distance – Closer proximity increases threat weight (0–2500 m with score 10, descending to score 4 for values greater than 5000).
4. Altitude – Lower altitudes receive higher scores (0–1500 m with score 10, descending to score 2 for values above 6000 m), aligning with the 8 km ceiling of the defence system.

As shown in Figure 3, the first five potential targets have a greater threat score.

4.3. Performance evaluation

Performance evaluation is conducted by testing and cross-referencing the results obtained using MCDMA TOPSIS model and the Baseline heuristic model. The weights of the TOPSIS algorithm are iterated from 0 to 1 (the sum of weights equals 1) to find the optimal parameters for the system in terms of the given criteria.

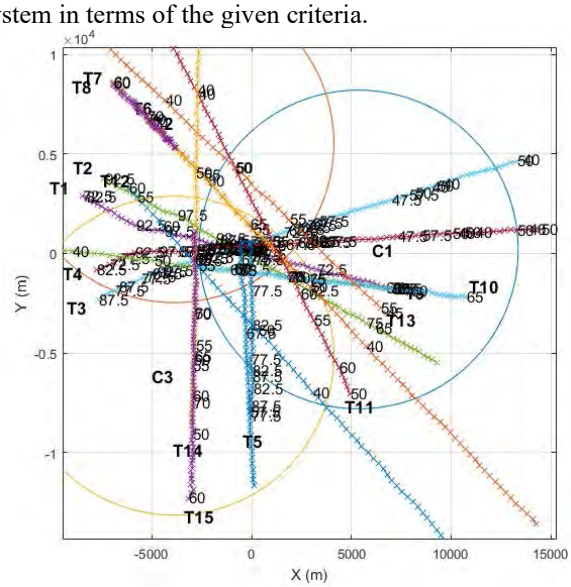


Figure 3 Tracks with calculated threat scores

The following cycle is the same for both algorithms:

1. The last update of tracks is used for simulation.
2. The engagement plan is calculated for all Track↔FU pairs.
3. The best pairs (depended of selected algorithm) are designated, and the destruction time is computed as:

$$t_e = t_c + (t_r + t_d) \times 1000 \qquad (1)$$

where:

t_e – total target engagement time
t_c – current scenario time
t_r – system reaction time
t_d – projectile-to-target flight (destruction) time

4. When the t_e time elapses, the track is destroyed, and the FU is freed for further use after its recovery time.
5. The test finishes when the condition is met - the most dangerous threats (T1-T5) have been destroyed.

The ITTE KPI for cross-referencing and performance assessment of both proposed algorithms is computed by following equation:

$$I_{te} = \int_{t_0}^{t_f} E_d(t)\, dt \qquad (2)$$

where:

I_{te} – ITTE
t_0 – scenario start time
t_f – scenario finish time
E_d – sum of threat scores of all remaining targets at moment t

Lower values indicate faster threat neutralisation and a reduced window of vulnerability.

5. RESULTS

The whole simulation is done using the Baseline algorithm first. As it considers only the distance between targets and FUs, it requires a long time to destroy high-intensity

manoeuvring targets (T1-T5).

For the TOPSIS-based MCDMA, extensive parameter sweeping is performed. Each of the four weights is varied from 0 to 1 with a step of 0.1, generating 10000 combinations. The complete simulation is executed for each combination.

Although the results for all 10000 combinations would be too extensive to present in full, three representative cases are selected and summarised in Table 2 to illustrate the impact of each criterion: one corresponding to the best performance, one to average performance, and one to the worst performance.

Based on the analysis of all simulated combinations, the following observations can be made:

1. Dominance of threat priority: Configurations emphasising calculated threat score ($w_1 \geq 0.4$) consistently yielded the lowest ITTE and fastest response times. The system successfully prioritised high-priority over low-priority targets.
2. Role of temporal criteria: A moderate weight on time to destruction ($w_2 = 0.1$) optimised kill timing without sacrificing threat prioritisation. The time-to-FU range criterion (w_3) showed negligible impact, as all targets entered engagement envelopes early in the scenario.
3. Degraded performance when threat is ignored: Configurations with $w_1 = 0$ produced the worst outcomes ($I_{te} > 31000$, $t_e \approx 69s$), confirming that distance-only or resource-balanced strategies fail under high-threat saturation.
4. Robustness and flexibility: Unlike the fixed Baseline algorithm, TOPSIS demonstrates adaptive behaviour. Suboptimal weight combinations gracefully degrade rather than fail, while optimal tuning delivers near-double the engagement speed and significantly reduced cumulative exposure.

Different combinations of weight factors, which demonstrate the impact of each factor, are given in Table 2.

Table 2 Results of different tuning

Algorithm	ITTE	TSET (s)
BASELINE	30307.5	61.5
TOPSIS (1)	21720	33
TOPSIS (2)	25627.5	42
TOPSIS (3)	31646.2	69

** TOPSIS (1) - $w_1 = 0.8$, $w_2 = 0.1$, $w_3 = 0$, $w_4 = 0.1$*
TOPSIS (2) - $w_1 = 0.3$, $w_2 = 0.2$, $w_3 = 0.2$, $w_4 = 0.3$
TOPSIS (3) - $w_1 = 0.1$, $w_2 = 0.7$, $w_3 = 0.1$, $w_4 = 0.1$

5.1. Comparative analysis

This section presents a quantitative and qualitative cross-referencing between the Baseline strategy and the proposed TOPSIS MCDMA. The analysis evaluates system performance based on the value of ITTE and TSET (in Table 3). The behaviour of each algorithm over the simulation duration is shown in Figure 4 and Figure 5.

Table 3 Comparative analysis summary

Metric	Baseline	TOPSIS (1)	Improvement
ITTE	30307.5	21720	28.33%
TSET	61.5 s	33 s	46.34 %

A review of the engagement logs reveals a fundamental difference in strategy between the two proposed models.

Baseline strategy: The Baseline algorithm engaged a total of 14 targets (T6, T9, T14, T13, T12, T3, T11, T1, T5, T10, T7, T15, T4, T2) as shown in Figure 4. While it neutralised a higher volume of targets, it required significantly more time to destroy the most critical targets, with the final engagement (T2) occurring at approximately 61.5 seconds.

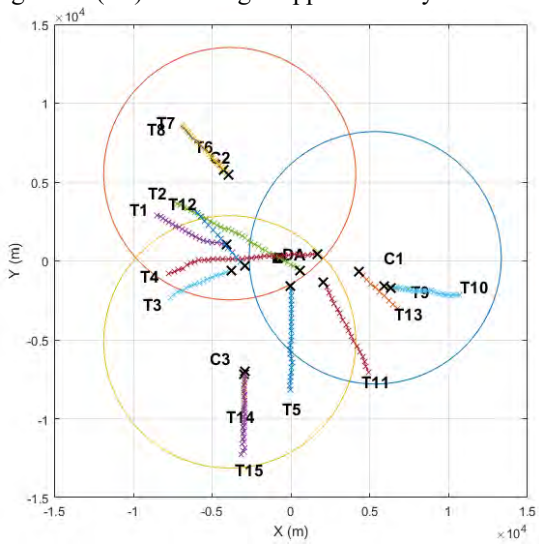


Figure 4 Baseline strategy results

TOPSIS strategy: The TOPSIS algorithm was configured to prioritise the most critical threats. It engaged a specific subset of 5 critical targets (T2, T3, T5, T4, T1) as shown in Figure 5.

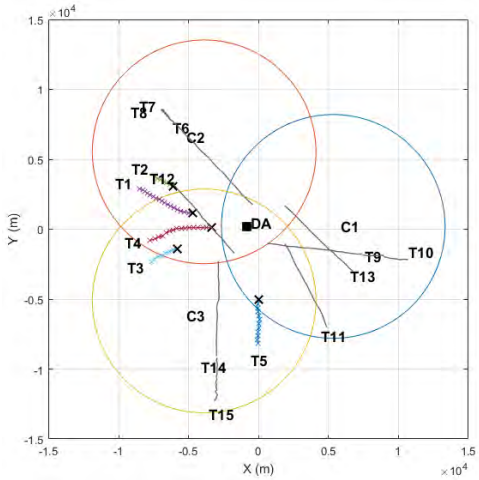


Figure 5 TOPSIS (1) strategy results

By focusing on high-priority targets first, the algorithm completed the engagement of the final critical target (T1) in approximately 33 seconds.

This demonstrates that while the Baseline approach is thorough, it is also inefficient regarding time criticality. The TOPSIS method provides a nearly 50% reduction in TSET. The second key observation illustrated in the figures is that the TOPSIS algorithm eliminates all high-priority threats at a considerable distance from the DA. In contrast, the Baseline algorithm failed to neutralise high-priority threats in time, allowing them to penetrate the DA's immediate vicinity.

Figure 6 illustrates the dynamic behaviour of the calculated threat score for both algorithms. Initially, the Baseline algorithm seems to lower the total threat score faster because it chooses the closest targets, even if they are not

the highest priority. In contrast, TOPSIS chooses the correct targets, but this process takes longer than simply destroying the nearest targets. This creates an illusion of efficiency in the Baseline algorithm, which masks a strategic failure.

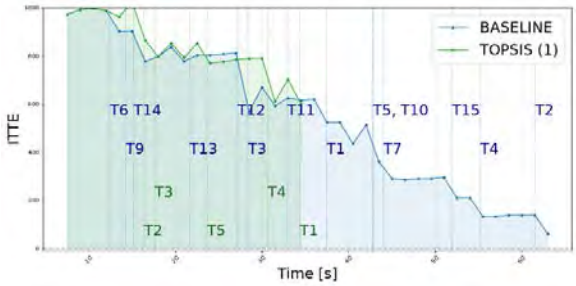


Figure 6 Threat score over time

6. CONCLUSION AND FUTURE WORK

In a modern warfare dynamic operational environment, automation of the decision-making process in GBAD and SHORAD systems is imperative. New threats such as loitering munitions, first-person view drones, and swarms provide means for the enemy to perform cheap and efficient strikes behind the first defensive line. The proposed approach enables efficient resource management of available FUs based on calculated threat score and MCDMA in a dynamic SHORAD operational environment. Results confirm the superiority of MCDMA-based TEWA against the heuristic "Greedy" strategy used to simulate quick human reaction. Considering FU constraints and ROE, as well as TOPSIS's dependency on a set of defined weighting factors for each individual criterion, further improvements toward Machine Learning and AI are possible. One possible improvement for future work is to define a large set of different scenarios and perform reinforcement learning based on finding an optimal set of weighting factors for MCDMA-based TEWA.

Acknowledgment

Ksenija Bokić, Ina Masnikosa and Tanja Lacmanović were supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, under Contract no. 451-03-33/2026-03/200034.

References

[1] AFSHAN, N., SHAOB A. K., ASAD W. M., *A real-time man-in-loop threat evaluation and resource assignment in defense,* Journal of the Operational Research Society (2017) 68(6), 725 - 738., 2017

[2] COSTA, I., COSTA, A., SANSEVERINO, A., GOMES, C., SANTOS M. *Bibliometric Studies on Multi-Criteria Decision Analysis (MCDA) Methods Applied in Military Problems,* Pesquisa Operacional. 42. 1 - 26., 2022

[3] KLINE, A., AHNER, D., HILL, R., *The Weapon-Target Assignment Problem,* Computers and Operations Research, Vol. 105, 226-236, 2019

[4] JOHANSSON, F., FALKMAN, G. (2009). *Performance Evaluation of TEWA Systems for Improved Decision Support*, Modeling Decisions for Artificial Intelligence, MDAI 2009. Lecture Notes in Computer Science, vol 5861. Springer, Berlin, Heidelberg, 2009

[5] ABERCROMBIE, B., COATES, G., MORRISS, S., UNDERWOOD, H., GRAVES, G., *A Systems Architecture for the Indirect Fire Engagement Process in a Dynamic Targeting Scenario*, A Regional Conference of the Society for Industrial and Systems Engineering, 57 – 62, 2020

[6] AIR FORCE DOCTRINE PUBLICATION 3-60, US Air Force, 2026

[7] Joint Air Power Competence Centre, *A Comprehensive Approach to Countering Unmanned Aircraft Systems,* Kalkar, Germany: JAPCC, 2020.

[8] JACO, N. R., VUUREN J. H., *Threat evaluation and weapon assignment decision support: A review of the state of the art,* ORiON. 23(2), 151-187, 2007

[9] GUERO-CONTRERAS, G., BALDERAS-DIAZ, S., GUTIERREZ-GELANO, L., FUENTES-LANDI G., DOMINGUES-JIMENEZ, J., *A Systematic Review of Artificial Intelligence in Air and Missile Defense*, International Conference on Information Systems Security and Privacy, 2026,

[10] FAUSTO, F.; GONZÁLEZ, A.; VALDIVIA-GONZÁLEZ, A.; García-Gutiérrez, V. *A Fuzzy Inference System for the Evaluation of Exploration and Exploitation Capabilities in Metaheuristic Optimization Algorithms,* Mathematics 2026, 14(14),

[11] WANG, L.; LIU, Z.G.; WANG, W., *Aerial Targets Threat Evaluation Method*, In Proceedings of the 2nd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2016), Xi'an, China, 2016

[12] PARADIS, S., BENASKEUR, A., OXENHAM, M. G., CUTLER P., *Threat evaluation and weapon allocation in network-centric warfare,* Proceedings of the Seventh International Conference on Information Fusion (FUSION), Stockholm, 1078–1085, 2005

[13] ZHANG, C., JIANLU, G., FEI, W., BOYUAN, C., CHUNSHI, F., LINGHUI, Y., ZHIWEN, W., *A dynamic parameters genetic algorithm for collaborative strike task allocation of unmanned aerial vehicle clusters towards heterogeneous targets,* Applied Soft Computing, Vol. 175, 2025

[14] ZHANG, X., X., JIANG, J., ZHOU, Y., *A Hybrid Approach of TOPSIS for Weapon Systems Selection*, Advanced Materials Research, vol. 850–851, 432–436, 2013

[15] *DOGAN, S. OZCAN, C., A hybrid AHP-TOPSIS framework for heterogeneous weapon allocation in air defense,* Defense Technology, Vol. 17, 1120 – 1132, 2021

[16] TEŠIĆ, D., BOŽANIĆ, D., PUŠKA, A.: *Application of MCDM Methods for Solving Military Problems: Literature Analysis from 2014 to 2024,* International Scientific Conference on Military Sciences VojNa, 2025

[17] TAHERDOOST, H., MADANCHIAN, M.: *A comprehensive guide to the TOPSIS method for multi-criteria decision making*, Sustainable Social Development.

[18] LÖTTER, DP., NIEUWOUNDT, I., VUUREN, JH.: *A multi-objective approach towards weapon assignment in a ground-based air defense environment*, ORiON. 29. 31., 2013.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SOFTWARE-DEFINED RADAR DIGITAL EXCITER, RECEIVER, AND SIGNAL PROCESSOR

MILOŠ D. JEVTIĆ
University of Belgrade, Institute „Mihailo Pupin", Belgrade, Serbia, e-mail milos.jevtic@pupin.rs

STEFAN STOJKOV
University of Belgrade, Institute „Mihailo Pupin", Belgrade, Serbia, e-mail stefan.stojkov@pupin.rs

NIKOLA LUKOVIĆ
Institute „Mihailo Pupin" – Computer Systems Ltd, Belgrade, Serbia, e-mail nikola.lukovic@pupin.rs

MILOŠ OTOVIĆ
Institute „Mihailo Pupin" – Computer Systems Ltd, Belgrade, Serbia, e-mail milos.otovic@pupin.rs

SRĐAN RADOVANOVIĆ
Institute „Mihailo Pupin" – Computer Systems Ltd, Belgrade, Serbia, e-mail srdjan.radovanovic@pupin.rs

INA MASNIKOSA
University of Belgrade, Institute „Mihailo Pupin", Belgrade, Serbia, e-mail ina.masnikosa@pupin.rs

DRAGAN NIKOLIĆ
Military Technical Institute, Belgrade, Serbia, e-mail nikolicdragansiki@gmail.com

DARKO PIJEVIĆ
Military Technical Institute, Belgrade, Serbia, e-mail darko.pijevic@gmail.com

DEJAN IVKOVIĆ
Military Technical Institute, Belgrade, Serbia, e-mail divkovic555@gmail.com

ZVONKO RADOSAVLJEVIĆ
Military Technical Institute, Belgrade, Serbia, e-mail zvonko.radosavljevic@gmail.com

Abstract: *We present a software-defined radar digital back-end (DBE) applicable in fully coherent radars with a power amplifier (PA) transmitter. DBE includes a digital exciter (DEX), receiver digital front-end (DFE), and radar signal processor (RSP). DEX performs direct digital synthesis of the transmit waveform and the local oscillator signal, along with synchronization signals. DFE accepts the intermediate-frequency signal and performs digital down-conversion, filtering, and decimation. Pulse compression, Doppler filtering, and constant false alarm rate detection are performed in the RSP. We use a commercial off-the-shelf, high-performance embedded computing platform to implement the DBE, with DEX and DFE implemented on a field-programmable gate array, and RSP implemented on a general-purpose processor. To test the DBE, we connect it with an existing PA-based analog front-end (AFE), antenna sub-system, and radar data processor. Field tests show that DBE enables the function of a fully coherent pulsed air-surveillance radar, capable of continuously tracking a multitude of aircraft, on both short and long distances, while successfully mitigating both clutter and noise interference. Although the field tests are performed only with very high frequency (VHF) band radar hardware, if connected with an appropriate AFE, DBE could be used in radars operating at higher frequencies.*

Keywords: *radar, receiver, signal processing, software defined, waveform synthesis.*

1. INTRODUCTION

Software-defined radar (SDR), with exciter, receiver, and signal processor realized digitally on reconfigurable hardware, is an established design approach with numerous implementations on commercial off-the-shelf (COTS) hardware [1]. This paper concerns the digital back-end (DBE) of such a radar, comprising: (1) a digital exciter (DEX) for direct digital synthesis (DDS) of the transmit waveform, local-oscillator (LO) signal, and trigger signals; (2) a receiver digital front-end (DFE) that digitizes the intermediate frequency (IF) echo and performs digital down-conversion (DDC), filtering, and decimation; and (3) a radar signal processor (RSP) for pulse compression (PC), Doppler filtering (DF), and detection. In a complete radar, the DBE operates with an

Corresponding author: M. D. Jevtić;
E-mail: milos.jevtic@pupin.rs
DOI: 10.5937/OTEH26082J

analog front-end (AFE) encompassing a power amplifier (PA) transmitter, and a radar data processor (RDP) for plot extraction and tracking.

Radars of this class transmit through a solid-state PA, whose modest peak power (commercial units reach the order of kilowatts [2]) is compensated by long pulses with intrapulse modulation and PC [3], [4]. Practical operation requires interleaving short and long pulses, transmitting at several pulse repetition frequencies (PRF), and suppressing clutter by moving target indication (MTI) or moving target detector (MTD) processing [3].

Reported SDR systems largely fall into two categories. General-purpose frameworks like GNU Radio with USRP-class hardware are used to build experimental processing chains, but operate at low power and short range, e.g., detecting a 0.1 m² drone out to about 150 m [5]. At the other end, dedicated DBEs implementing the full pulse-Doppler chain (DDC, PC, MTD, and constant-false-alarm-rate (CFAR) detection) have been realized on custom field-programmable gate array (FPGA) boards for specific radars [6]. The DBE described here occupies a different point in the SDR space: it is built on a COTS high-performance embedded computing (HPEC) platform, partitioning the DEX and DFE onto an FPGA and the RSP onto a general-purpose processor (GPP), aiming at applicability in fully coherent PA-based radars.

The contributions of this paper are the following: (1) an integrated, COTS-based radar DBE that combines the DEX, DFE, and RSP; (2) a hardware/software partitioning that maps the DEX and DFE functions onto an FPGA and the RSP onto a GPP; (3) support for complex waveforms, including interleaved short and long pulses and multiple PRFs, together with MTI or MTD clutter suppression; (4) a demonstration that the DBE, combined with a PA-based AFE and an antenna, constitutes a functioning fully coherent very high frequency (VHF) air-surveillance radar, validated in the field against live air traffic, with applicability to higher frequency bands given a suitable AFE.

2. REQUIREMENTS AND CONSTRAINTS

The operating concept from Section 1 imposes the following requirements on the DBE:

- it shall provide fully coherent operation;
- it shall operate in real time;
- the transmit waveform shall support configurable intrapulse modulation;
- it shall support interleaving of two pulses, short and long, the short pulse restoring the short-range coverage lost to eclipsing during the long pulse [3];
- the carrier frequency, the pulse duration, and the pulse repetition interval (PRI) shall be configurable;
- it shall support pulse-to-pulse carrier-frequency and PRI agility;
- it shall mitigate clutter and noise interference, with multiple-PRF operation used to circumvent the effects of the blind Doppler intervals introduced by clutter suppression [3].

3. SYSTEM ARCHITECTURE

The DBE comprises three logically distinct subsystems: DEX, DFE, and RSP. The DEX and DFE execute as deterministic hardware pipelines on the FPGA, while the RSP runs as software on the GPP. Both processing domains are hosted on a single COTS HPEC platform, which provides the shared backplane, deterministic synchronization, and high-bandwidth data transport.

The block diagram of DBE is shown in Figure 1. The transmit signal flows from the DEX through a digital-to-analog (D/A) converter to the AFE for power amplification. The DEX generates triggers that coordinate the AFE transmit/receive switch and signal the start of reception. Echo returns are collected by the antenna sub-system, amplified and down-converted by the AFE to the analog IF (AIF), and presented to the DFE for digitization. The DFE extracts the complex envelope and applies decimation. The decimated complex envelope samples are organized into frames, together with the azimuth value at the moment of the originating pulse transmission and pulse metadata. Frames are transferred from the FPGA to GPP memory via direct memory access (DMA) and delivered to the RSP through a callback interface. After processing, the RSP produces a real-valued radar video signal, framed with azimuth, and passes it to an external RDP for downstream plot extraction and tracking [3, Fig. 1.2].

The DBE's COTS HPEC platform is based on the NI FlexRIO system. It comprises a transceiver module providing analog-to-digital (A/D) and D/A conversion, an FPGA module, a real-time controller based on a GPP, and a shared chassis. FlexRIO modules have been employed in radar applications before [7].

The partitioning of the DBE follows the principle: sample-synchronous processing on the FPGA, frame-based processing on the GPP. The DEX and DFE form streaming, sample-synchronous pipelines with strict latency bounds. The RSP's algorithms are compute-intensive but tolerate the latency of frame assembly, and benefit from the numerical precision and vectorization capabilities of a modern GPP, the flexibility of software reconfiguration and the easier integration with RDP.

4. SUBSYSTEM DESIGN

4.1. Digital Exciter (DEX)

The DEX is built on the DDS principle [4, Ch. 6]. Its block diagram is shown in Figure 2. The transmit waveform is defined as a pre-computed complex envelope of a pulse, stored in FPGA memory. The pulse duration and all intrapulse modulation parameters of the waveform are set from the configuration transferred from the GPP. The complex envelope is up-sampled by an interpolation chain comprising a multiplierless cascaded integrator-comb (CIC) filter, followed by a finite impulse response (FIR) filter that compensates the CIC amplitude response and provides precise anti-imaging filtering [8]. Then, the complex envelope is translated to the digital IF (DIF) by complex multiplication with a complex sinusoid from a

numerically controlled oscillator (NCO). The DIF is configurable, allowing the carrier frequency to be adjusted. The digital complex pulse at the DIF is interpolated to the output sampling rate by a digital interpolator inside the D/A converter on-board the transceiver module. The D/A converter then applies a digital complex modulator to up-convert the signal from the DIF to the carrier frequency. The I component of the resulting signal is D/A-converted, and the output is cleaned by on-board and external analog low-pass filters.

The LO signal is synthesized independently as a complex sinusoid at a fixed DIF value by a separate NCO on the FPGA. The output is interpolated, up-converted, and D/A-converted through a separate D/A channel on the transceiver module, producing the analog sinusoid that drives the AFE's analog down-conversion mixer. Both the waveform DIF mixer and the LO NCO operate within the same FPGA clock domain. The per-pulse phase reset of both NCOs ensures that each transmitted pulse has a well-defined, repeatable phase relationship with the LO, which is required for coherent reception and DF.

The trigger pulses that mark the instant of each pulse transmission are output from the FPGA through the transceiver's digital I/O lines. Triggers are derived from the same FPGA clock that drives waveform synthesis. Trigger timing and DIF are configurable per-pulse from a preset sequence transferred from the GPP.

4.2. Receiver Digital Front-End (DFE)

The DFE (Figure 3) comprises three stages: (1) A/D conversion of the AIF echo signal; (2) DDC from the AIF to the complex baseband; and (3) a filtering and decimation chain that reduces the sampling rate while providing spectral selectivity. The A/D conversion is performed on the transceiver module, while the DDC and decimation execute on the FPGA.

The AIF signal is digitized by the A/D converter of the transceiver module. Collection of the converted samples is synchronized with the trigger routed from the DEX.

The complex envelope of the digitized real-valued AIF signal is extracted by complex mixing with a digital complex sinusoid [8]. The sinusoid is generated by an NCO set to the AIF, with its initial phase reset on each trigger edge (essential for coherence across pulses).

The complex envelope is decimated using a cascaded pair of FIR decimators, each performing low-pass filtering followed by down-sampling by an integer factor [8].

Assembled complex envelope frames are written to a first-in-first-out buffer on the FPGA, from which they are read by the GPP via DMA and delivered to the RSP.

4.3. Radar Signal Processor (RSP)

The RSP processing chain (Figure 4) comprises five stages: (1) PC; (2) DF; (3) envelope detection; (4) CFAR detection; and (5) frame combining. These stages are applied sequentially, with the DF stage being the only point of branching between MTI and MTD modes.

PC resolves targets in range by matched filtering the received echo against the transmitted pulse template. It is implemented via multiplication in the discrete Fourier transform (DFT) domain [3]. A Taylor window is applied to the matched-filter coefficients for sidelobe control.

To suppress clutter, DF is performed along the slow-time axis (across successive pulses at each range gate), with two modes available. The first mode is MTI [3], implemented as a FIR filter, with configurable order, applied only within a configurable range window.

As an alternative, MTD processor [3], [9] is available. It transforms a coherent processing interval (CPI)-length sequence of complex input frames into a Doppler-resolved output via two parallel channels. The zero-velocity channel retains near-DC returns: a low-pass integrator accumulates energy across the CPI, followed by magnitude computation and an adaptive range-azimuth clutter map update via exponential averaging. The clutter map based detector compares the magnitudes against the learned clutter levels to produce a detection mask for the zero-Doppler bin. The non-zero Doppler channels begin with a high-pass FIR filter. A Taylor-windowed DFT bank then resolves returns across Doppler frequency for all range gates simultaneously. After magnitude computation per Doppler channel, a sliding-window cell-averaging CFAR (CA-CFAR) [3] detector is applied in the range domain. The outputs of all channels pass through a pulse-width discriminator and a primitive observation extractor, followed by binary M-of-N integration across successive CPIs. The video synthesizer constructs complex output frames from the observations, making MTD a drop-in replacement for MTI.

After DF, envelope detection converts the complex-valued signal to a real-valued magnitude.

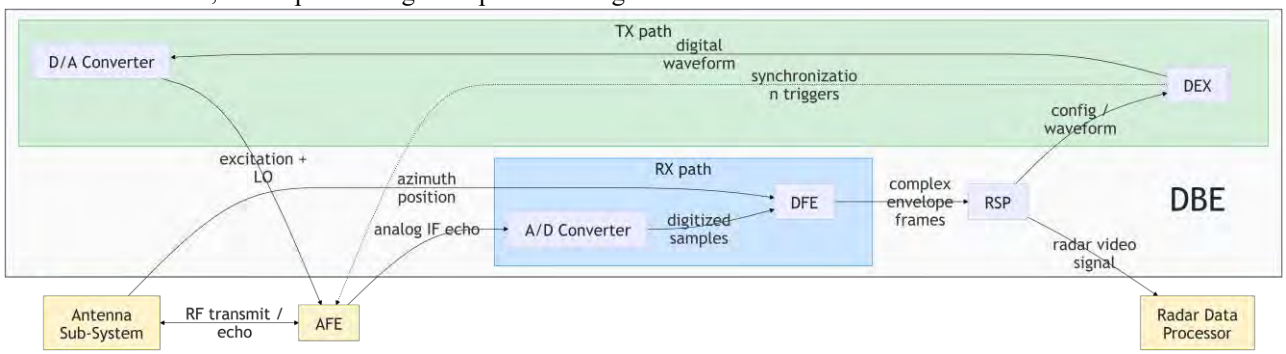


Figure 1. Top-level block diagram: antenna sub-system — AFE — DBE (DEX / DFE / RSP) — RDP.

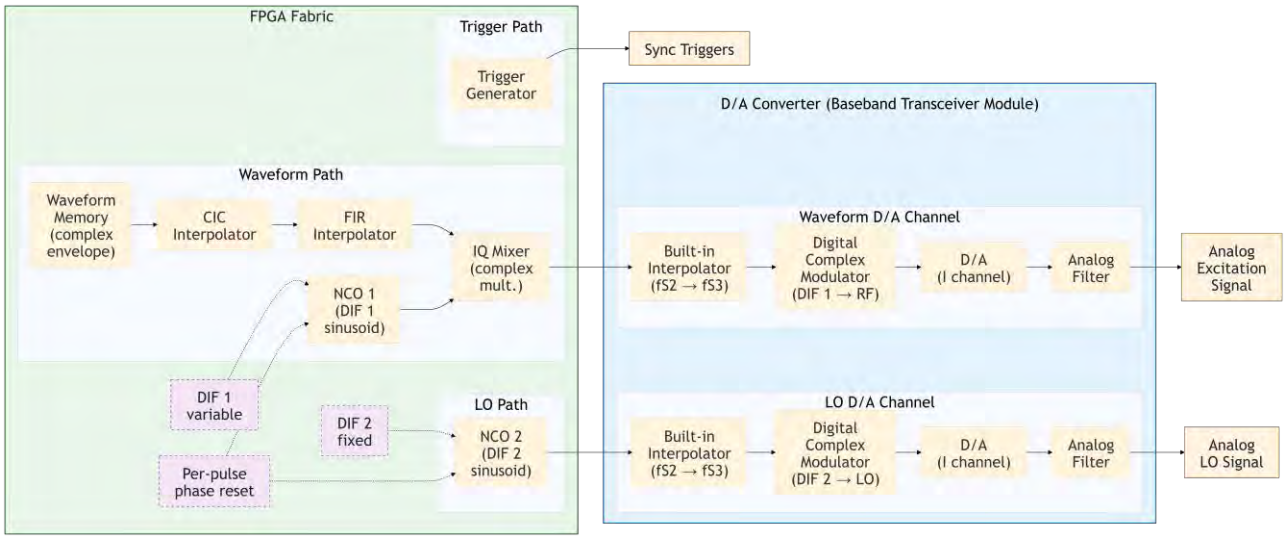


Figure 2. Block diagram of the digital exciter (DEX).

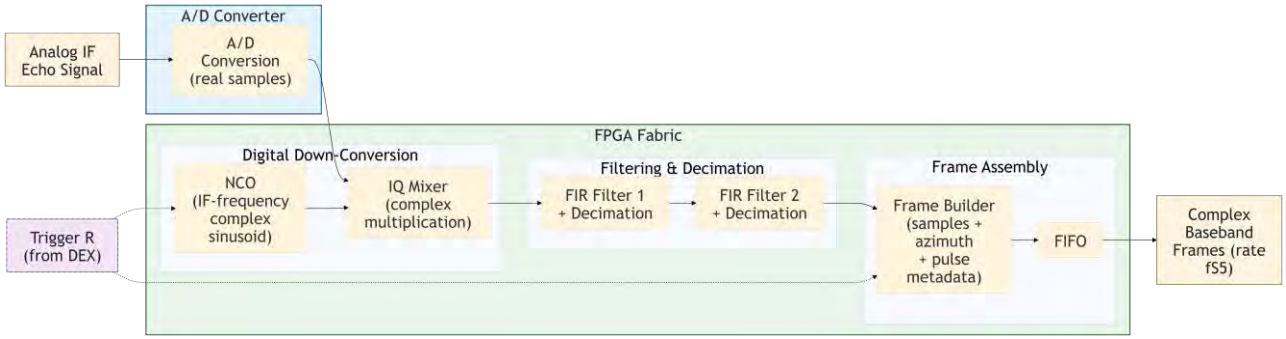


Figure 3. Signal chain of the receiver digital front-end (DFE).

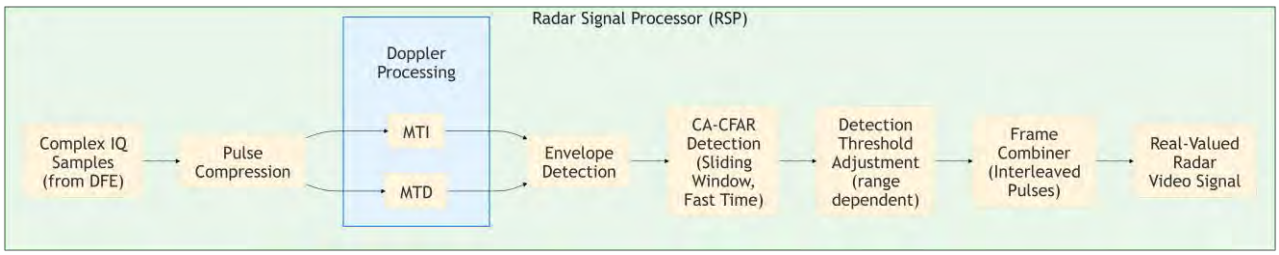


Figure 4. Radar signal processor (RSP) processing chain.

The CFAR detector computes an adaptive threshold from the local noise-plus-clutter environment using a sliding window along the fast-time (range) axis. When the MTI path is active, the detector uses CA-CFAR, with configurable window width, guard-cell count, CFAR constant, and bias. When the MTD path is active, the outer detection stage is bypassed, since the MTD already performs CFAR detection. An optional detection threshold adjustment provides range-dependent scaling of the threshold.

The frame combiner is active only in the interleaved short-and-long-pulse mode. It concatenates the near-range samples from the short-pulse output and the far-range samples from the long-pulse output. The final output is a real-valued radar video signal, accompanied by the azimuth value at transmission time.

5. FIELD TEST AND RESULTS

The DBE was integrated with an existing PA-based AFE, a rotating antenna subsystem designed for VHF-band operation, and a RDP running on a GPP. The RDP comprises a plot extractor [10], a tracker [11], and a radar display application that visualizes radar video, tracks, and annotation overlays [12].

Evaluation was carried out using two complementary operational modes. In live field mode, the DBE was connected directly to the AFE and antenna, with echo signals from real airborne targets digitized and processed in real time. The complex envelope data entering the RSP was recorded digitally by the RSP. In replay mode, recorded data is injected into the same data entry path used by live hardware, ensuring the same input signal and processing pipeline as in the live field mode.

From the available field recordings, four situational windows were selected for evaluation. These “scenarios” were opportunistic observations captured from live air traffic during the recording sessions. After preliminary analysis, stationary clutter fields were mapped by correlating radar returns with terrain data. To prevent the

formation of tracks on dense clutter returns, a circular forbidden track-initiation zone was configured around the radar site. The four scenarios are described below.

The objective of scenario 1 was to demonstrate sustained tracking of a large number of simultaneous airborne targets. Figure 5 shows the result. Continuous tracks are maintained across the surveillance volume, confirming that the DBE's processing pipeline supports simultaneous tracking of multiple targets under realistic clutter conditions. The result demonstrates that the DBE, when combined with the antenna sub-system, the AFE and the RDP, satisfies the requirement for air-surveillance operation over a wide coverage area.

Scenario 2 compared MTI and MTD processing in the context of tracking far-range targets. The recorded signal was re-processed with two different configurations: MTI and MTD, both evaluated against the same set of targets. Figure 6 presents the two processing variants side by side. In both MTI (top) and MTD (bottom) configurations, the far-range target at ~245 km is tracked successfully. Both processing variants are capable of tracking targets at long range on this dataset, demonstrating operational flexibility in the choice of DF strategy.

The 3rd scenario evaluated tracking performance at short range, where the eclipsing caused by long-pulse transmission can leave a blind zone if not compensated. As shown in Figure 7, targets at 9 km (blue) and 15 km (yellow) are tracked consistently. Critically, the interleaved short pulse restores coverage in the near-range zone that would otherwise be eclipsed [13].

The 4th scenario quantified the impact of DF on track quality when targets traverse a clutter field originating from a known mountain. The tracks of three targets passing through the clutter field were evaluated under three processing configurations: no DF, MTI, and MTD (Figure 8). In the left panel (no DF), strong clutter returns generate numerous false detections. Multiple spurious tracks appear in the clutter field, and the legitimate track histories for two of the three targets are corrupted by clutter-induced detections. In the center panel (MTI), the stationary clutter is suppressed, and the target tracks are clean and continuous. In the right panel (MTD), the target tracks are similarly clean and continuous. The before-and-after comparison shows that both MTI and MTD achieve an adequate level of clutter suppression on this dataset.

6. CONCLUSION

This paper presented a SDR DBE that integrates a DEX, a DFE, and a RSP on a COTS HPEC platform. The DEX and DFE were mapped to a FPGA as deterministic, sample-synchronous hardware pipelines, while the RSP runs as software on a GPP. This FPGA/GPP partitioning exploits the complementary strengths of each domain.

The DEX employs DDS with CIC + FIR interpolation and per-pulse NCO phase reset to guarantee full coherence across the pulse train. The DFE performs DDC and two-stage FIR decimation. The RSP chains frequency-domain PC, DF (either MTI or MTD), envelope detection, and CFAR detection into a single configurable pipeline.

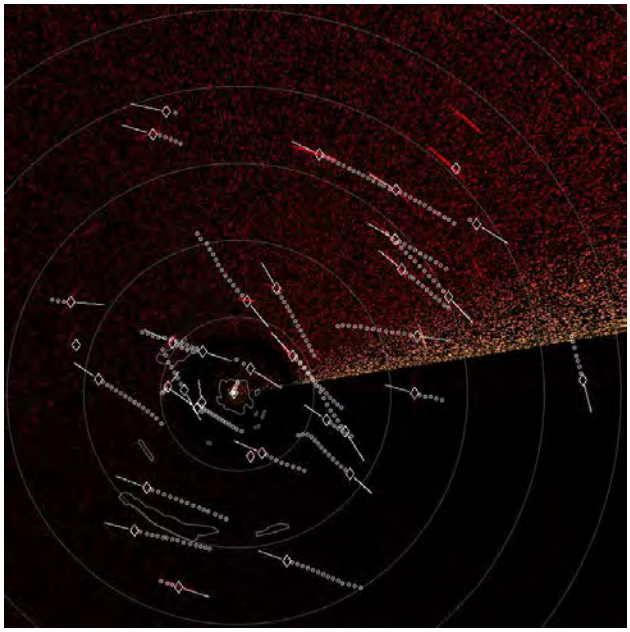


Figure 5. Multi-target tracking. Radar video (red), tracks (white), clutter fields (gray polygons).

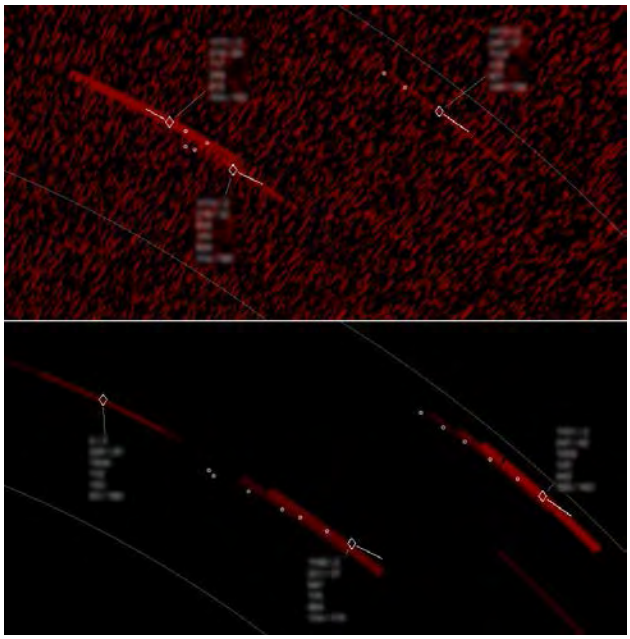


Figure 6. Far-range tracking — MTI (top) versus MTD (bottom). Radar video (red), tracks (white).

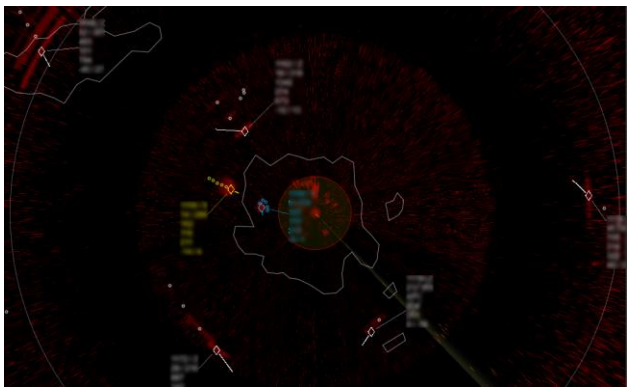


Figure 7. Near-range tracking. Radar video (red), tracks (color-coded), clutter fields (gray polygons).

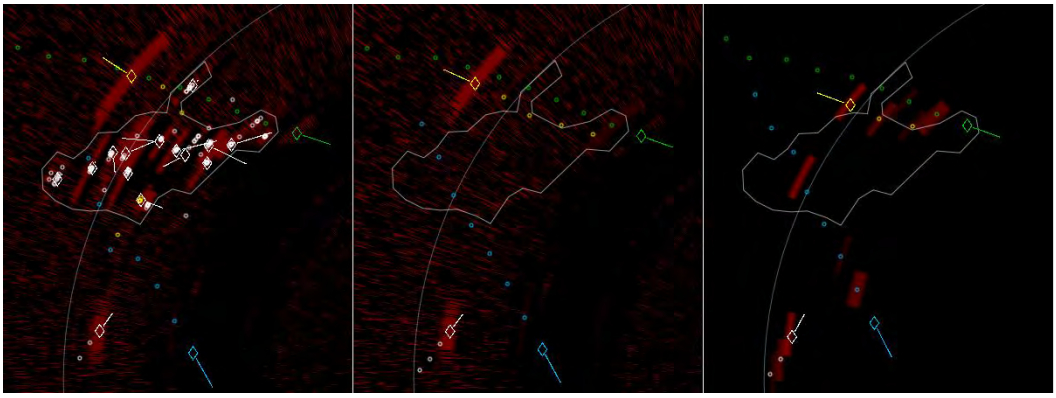


Figure 8. Clutter-field tracking — Doppler filtering disabled (left), MTI (center), MTD (right). Radar video (red), tracks (color-coded), clutter field (gray polygon).

The DBE supports complex waveforms including interleaved short and long pulses, and pulse-to-pulse carrier-frequency and PRI agility.

Field tests against live air traffic, carried out with the DBE integrated into a proof-of-concept PA-based VHF air-surveillance radar, demonstrated continuous tracking of multiple simultaneous airborne targets across the wide area, including far-range target at ~245 km and near-range target at ~9 km. Both MTI and MTD processors were shown to provide effective clutter mitigation.

Although validated with VHF-band hardware, the DBE's signal-processing architecture is frequency-agnostic. Given an appropriate AFE and antenna sub-system, the same DEX, DFE, and RSP pipeline can be applied to radars operating at higher frequency bands, consistent with the system's software-defined design philosophy.

Future work could include deploying the DBE in higher frequency bands to validate its scalability, extending the RSP with additional CFAR and Doppler-processing algorithms, and extending the RSP's interface to RDP.

Acknowledgment

Miloš D. Jevtić, Stefan Stojkov and Ina Masnikosa were supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, under Contract no. 451-03-33/2026-03/200034.

References

[1] S. Flandermeyer, R. Mattingly, and J. Metcalf, "gr-plasma: A New GNU Radio-based Tool for Software-defined Radar," in *Proceedings of the GNU Radio Conference*, Sep. 2022. Accessed: Jun. 19, 2026. [Online]. Available: https://pubs.gnuradio.org/index.php/grcon/article/view/121

[2] Empower RF Systems, "Solid State Broadband High Power Amplifier 2244: 1000 – 1400MHz, 8 kW Pulsed." Empower RF Systems, Inglewood, CA, 2022. Accessed: Jun. 19, 2026. [Online]. Available: https://www.empowerrf.com/datasheet/Empower_RF_Amplifier_2244.pdf

[3] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, NY, USA: McGraw-Hill, 2014.

[4] M. I. Skolnik, Ed., *Radar Handbook*, 3rd ed. New York, NY, USA: McGraw-Hill, 2008.

[5] T. W. Mathumo, T. G. Swart, and R. W. Focke, "Implementation of a GNU radio and python FMCW radar toolkit," in *2017 IEEE AFRICON*, Sep. 2017, pp. 585–590. doi: 10.1109/AFRCON.2017.8095547.

[6] M. Yang, J. Yang, Y. Hou, and C. Jin, "Implementation architecture of signal processing in pulse Doppler radar system based on FPGA," *J. Eng.*, vol. 2019, no. 21, pp. 7335–7338, 2019, doi: 10.1049/joe.2019.0644.

[7] R. Abbasi *et al.*, "Telescope Array Radar (TARA) observatory for Ultra-High Energy Cosmic Rays," *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.*, vol. 767, pp. 322–338, Dec. 2014, doi: 10.1016/j.nima.2014.08.015.

[8] R. G. Lyons, *Understanding Digital Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2011.

[9] L. Cartledge and R. M. O'Donnell, "Description and Performance Evaluation of the Moving Target Detector," MIT Lincoln Laboratory, Lexington, MA, USA, Project Report ATC-69, Mar. 1977.

[10] M. Jevtić and M. Stamatović, "Radar Data Processing and Visualization on Desktop Platforms," in *Proc. 17th Telecommunications Forum (TELFOR)*, Belgrade, Serbia, Nov. 2009, pp. 1315–1318.

[11] M. Jevtić, K. Marković, D. Mikluc, B. Mišić, and T. Pajić, "Real-time Implementation of Target Tracking System for Air Surveillance Radar Applications," in *Proc. 11th Int. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*, Niš, Serbia, Oct. 2013, pp. 557–560. doi: 10.1109/TELSKS.2013.6704440.

[12] M. Stamatović, M. Jevtić, U. Kisić, M. Tatarević, T. Pajić, and K. Marković, "Modern Air Situation Picture Display for Air Surveillance Radar Applications," *Telfor J.*, vol. 5, no. 1, pp. 54–59, 2013.

[13] A. Stocker, "Dwelltime Minimization for Radar Target Detection in the Clear," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 130–138, Jan. 1987, doi: 10.1109/TAES.1987.313342.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PREVENTIVE EM MEASUREMENTS: COST-EFFECTIVE SOLUTION FOR AWARENESS OF EMERGING TEMPEST THREATS

DRAGAN KONDIĆ
Applied Mathematics and Electronics Centre, Belgrade, dragan.kondic@vs.rs

BORKO DJAKOVIĆ
Applied Mathematics and Electronics Centre, Belgrade, borko.djakovic@vs.rs

MILENA GRDOVIĆ
Applied Mathematics and Electronics Centre, Belgrade, milena.grdovic@vs.rs

NENAD STEFANOVIĆ
Applied Mathematics and Electronics Centre, Belgrade, nenad.stefanovic@vs.rs

Abstract: *Emerging TEMPEST threats are gaining relevance due to increasing availablity of open platforms and publicly software that facilitate electromagnetic (EM) side-channel experimentation. Most commercial off-the-shelf and industrial-grade devices, as potential targets, do not address these vulnerabilities, while awareness of associated risks remains limited. Conversely, procurement of TEMPEST resistant devices, structures and facilities often requires heavy budget strains. Preventive EM measurements of protected entities are therefore essential for assesing exposure and identifying vulnerabilities. However, such measurements can be costly and resource-intensive, and unlike electromagnetic compatibility testing, where pre-compliance frameworks exist to reduce risk and cost, TEMPEST evaluation currently lacks formal intermediate assesment stages. As a result, organizations face significant uncertainty and potential expense if measurement processes are not properly structured. This paper proposes a hybrid methodology that integrates engineering-level pre-compliance EMSEC measurements with physical security measures, including shielding effectiveness (SE) assesment and EM security zoning. The approach leverages publicly available ITU-T Recommendations K.81, K.84, and K.87 to provide a systematic, repeatable and extensible framework for early-stage risk assesment and mitigation. By combining measurement-driven evaluation with spatially structured protective measures, the methodology bridges the gap between open-standard practices and classified TEMPEST requirements.*

Keywords: *side-channels, EMSEC, TEMPEST, zoning, shielding, profiling.*

1. INTRODUCTION

In an era where spectrum dominance dictates cross-domain security, mitigating data leakage via compromising electromagnetic emanations (CE) has become a critical priority at the convergence of cyber and physical warfare, Figure 1. Traditionally, exploiting such vulnerabilities required highly specialized equipment and expertise, limiting practical attacks to well-funded state actors or advanced laboratories. However, the threat landscape is rapidly evolving due to the increasing availability of open platforms and publicly accessible software tools that facilitate EM side-channel experimentation [1]. Recent comprehensive research surveys on EM information leakage [2] is followed by an on-line available tool to conduct some tests. Beyond phreaking video signals from monitors, recently published report describes similar leakage from cameras as well, including free tools for signal reconstruction [3]. Other give more general degrees of freedom creating more specific reconstruction templates [4]. Other closed-tool publications report on succsessfull recovering of other classes of emanations such as from touchscreens, smartphones [5], serial cables and broad list of other physical asserts. Furthermore, novel class of available threats is usage of sending informations over unintended emissions from electronic devices with activity of simple unprivileged code [6], showing increased risk that CE carry. However, in order to procure TEMPEST resistant devices, structures and facilities, expose high costs if not structured properly. Finally, to challenge the common assumption that compromising emanations are confined to short-range effects, recent studies on so-called *"screaming channels"* have demonstrated that sensitive information can couple onto RF carriers and propagate over significantly longer distances [7].

In this context, the paper proposes a hybrid EMSEC pre-compliance framework derived from publicly available ITU-T recommendations and extended through analytical propagation and probabilistic compromise models.

The proposed approach combines intrusion-zone-oriented assessment, shielding evaluation, and receiver-oriented measurement interpretation to support systematic early-stage EMSEC evaluation. The objective of proposed framework is not to replace formal TEMPEST evaluation, but to provide a structured engineering methodology for preliminary EMSEC assestment and mitigation planning.

Corresponding author: N. Stefanović;
E-mail: nenad.stefanovic@vs.rs

DOI: 10.5937/OTEH26081K

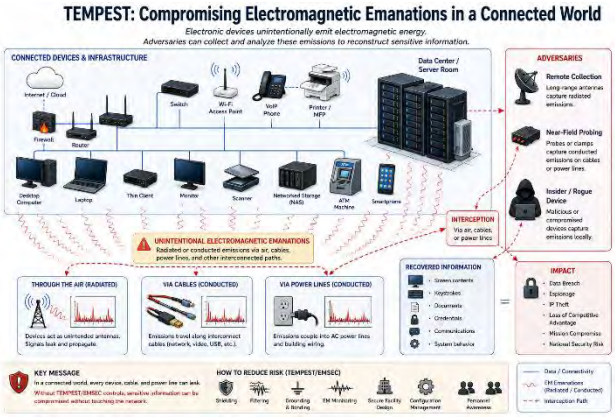


Figure 1. Overall description of CE threats

The paper contributions are:

- Interpretation of ITU-T K.84 through a propagation-oriented EMSEC threat model.
- Introduction of an engineering-level EMSEC pre-compliance workflow inspired by EMC pre-compliance practices.
- Development of a probabilistic compromise model linking emanation level, shielding, propagation distance and receiver capability.
- Extension of intrusion zones from geometric boundaries into probability-based exposure regions.
- Demonstration of measurement-driven mitigation optimization through illustrative examples.

2. EMSEC THREAT MODEL

2.1. Compromising Emanation Threat Landscape

Electromagnetic security assessment defined in ITU-T K.84 is fundamentally based on the relationship between the spatial intrusion threat model and the measurable compromising emanations generated by information processing equipment, Figure 2. Unlike conventional EMC testing, the purpose of EMSEC measurements is not regulatory compliance but evaluation of information reconstruction feasibility at different intrusion zones. The recommendation implicitly defines a spatially dependent threat model where the electromagnetic field strength at the attacker location depends on the original emanation level, equipment countermeasures, building shielding, propagation distance and receiver sensitivity.

Consequently, the measurement methodology must be adapted to the assumed intrusion zone and attacker capabilities.

The measurement methodology is directly linked to the assumed intrusion zone, table 1. Lower intrusion zones correspond to shorter attacker distances and therefore require evaluation of higher received field strengths and wider receiver capabilities.

Conversely, higher intrusion zones are dominated by propagation attenuation and shielding effects, making narrow-band spectral detection more suitable for evaluating residual leakage components.

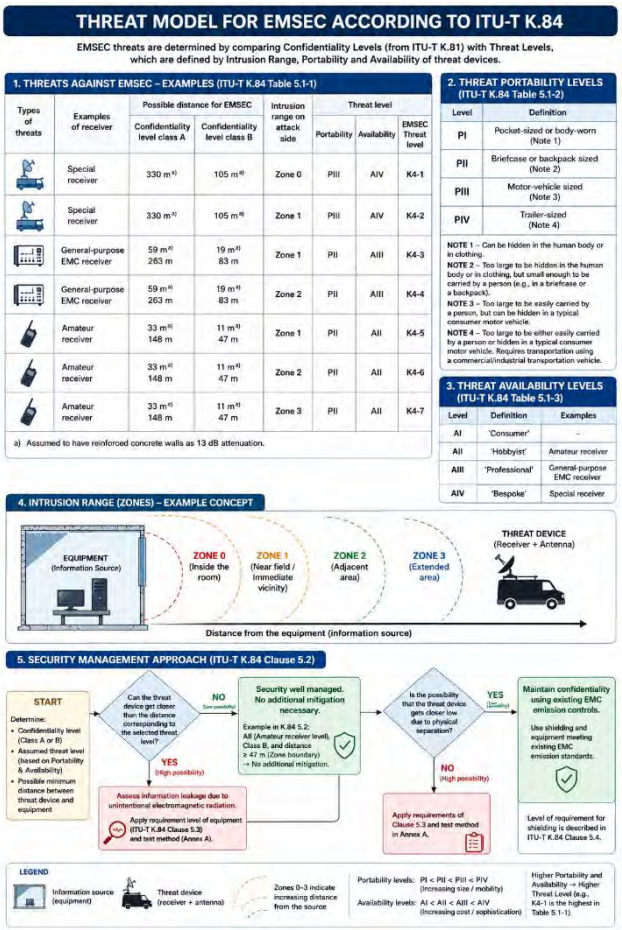


Figure 2. Modeling threats acording to ITU-T K.84

ITU-T K.84 implicitly separates EMSEC evaluation into two complementary dimensions:

1. Spatial threat evaluation through intrusion zones and propagation attenuation;
2. Spectral leakage evaluation through wideband and narrow-band measurement techniques.

Table 1. Intrusion EMSEC zones

Intrusion Zone	Dominant threat	Preferred measurement meth.	Objective
Zone 0	Direct reconstruction	*Wideband*	*Image reconstruction capability*
Zone 1	Adjacent room leakage	*Wideband + Narrowband*	*Reconstruction + spectral analysis*
Zone 2	Boundary leakage	*Narrowband*	*Leakage frequency identification*
Zone 3	Extended perimeter	*Narrowband*	*Residual spectral detectability*

The combination of these approaches forms a practical framework for estimating the feasibility of electromagnetic information compromise under realistic operational conditions.

Recommendation identifies several possible mechanisms for compromising electromagnetic emanation acquisition, including free-space radiated interception using antennas, conductive coupling through metallic structures and

leakage acquisition through power-supply lines.

2.2 Wideband and Narrow-Band EMSEC Measurements

ITU-T K.84 practically introduces two different measurement paradigms, table 2.

Table 2. EMSEC measurement approaches

Method	Purpose	Security interpretation
Wideband	Reconstruction capability	*Can attacker reconstruct information?*
Narrowband	Spectral leakage detection	*Which frequencies carry compromising information?*

The wideband method attempts to emulate the operational conditions of a realistic compromising emanation receiver capable of reconstructing video information. The measurement process approximates the operational capability of advanced EMSEC threat receivers described in higher threat classes. The wideband method preserves the energy distribution of compromising emanations over a broad frequency interval, enabling estimation of reconstructed information quality rather than isolated spectral components.

The narrow-band method focuses on identification of discrete spectral components correlated with assumed information patterns. Frequencies exhibiting significant amplitude differences between static and time-varying patterns are identified as potential compromising emanation carriers. Consequently, the narrow-band method represents a frequency-domain leakage localization technique rather than a direct reconstruction method.

The minimum electromagnetic field level [dBμV/m], necessary for successful information reconstruction is expressed as:

$$E_{min} = V_{min} + SN + A_f \quad (1)$$

where: V_{min} [dBμV] denotes minimum detectable receiver voltage, SN [dB] denotes the signal-to-noise ratio required for reconstruction, and $A_f[dB]$ denotes the antenna factor. This formulation directly links electromagnetic propagation to information reconstruction feasibility. Consequently, successful EMSEC interception requires careful optimization between receiver bandwidth, amplifier gain, noise figure and spectral filtering.

The recommendation categorizes interception receivers into three operational classes: amateur receivers, professional EMC receivers and specialized EMSEC receivers. These categories correspond to progressively improved sensitivity, signal processing capability, bandwidth and reconstruction performance. The interception problem may therefore be interpreted as a competition between propagation attenuation, shielding attenuation and receiver sensitivity enhancement. This interpretation directly motivates the use of probabilistic compromise models, described later in section 4.2.

3. EMSEC PRE-COMPLIANCE FRAMEWORK

3.1 Cost-Effective Measurement-Driven EMSEC Assessment

EMSEC protection mechanisms such as facility shielding, filtered infrastructure, protected zoning, and TEMPEST-certified equipment may introduce substantial financial and operational costs if implemented without prior leakage characterization. Unlike conventional EMC mitigation, where compliance limits are standardized and well understood, EMSEC protection depends strongly on equipment-specific compromising emanation profiles, installation geometry, building materials, cable routing and attacker positioning relative to the protected asset. Consequently, the effectiveness of countermeasures cannot be reliably estimated without measurement-driven assessment.

In practical environments, compromising emanations are rarely uniformly distributed across the spectrum or spatial environment. Certain frequency bands may exhibit significantly elevated leakage levels due to internal clock harmonics, display synchronization frequencies, cable resonances, power conversion stages or unintended antenna structures formed by connected conductors.
Similarly, electromagnetic leakage paths are highly dependent on physical installation conditions. Openings, windows, ventilation structures, cable penetrations, and conductive building elements may create localized reductions in shielding effectiveness (SE), enabling disproportionate propagation toward specific intrusion zones.

As a result, implementation of generalized shielding measures without prior measurements may produce excessive shielding costs, unnecessary infrastructure modifications, overestimation of threat exposure, or insufficient protection against dominant leakage mechanisms.

The measurement methodology defined in ITU-T K.84 therefore serves not only as a security evaluation procedure, but also as an engineering optimization mechanism for determining dominant leakage frequencies, effective propagation directions, critical intrusion zones and required attenuation margins. Wideband measurements enable estimation of practical information reconstruction feasibility, while narrow-band measurements identify dominant spectral leakage components. Their combined application supports prioritization of mitigation measures according to measurable compromise potential rather than purely theoretical assumptions.

3.2 Facility-Oriented Workflow

The proposed analytical propagation framework further extends this concept by enabling preliminary estimation of compromise distance, shielding requirements and intrusion-zone exposure before deployment of expensive TEMPEST-oriented infrastructure modifications.
The conceptual facility-oriented assessment shown in Figure 3 illustrates how EMSEC pre-compliance

evaluation may be integrated into realistic operational environments. Instead of treating compromising emanations as purely device-level phenomena, the proposed approach considers the combined influence of equipment leakage characteristics, building geometry, SE, intrusion-zone proximity and propagation-oriented measurement positioning.

Wideband and narrow-band measurements are strategically positioned according to expected attacker locations and dominant propagation paths. The resulting measurements are then incorporated into the proposed analytical propagation framework to estimate residual compromise probability across the protected facility. This interpretation enables targeted deployment of shielding and filtering countermeasures only where measurements indicate elevated compromise potential, thereby reducing unnecessary infrastructure hardening and minimizing over-engineering costs.

3.3 Proposed Pre-Compliance Methodology

The proposed framework may be interpreted as a staged EMSEC pre-compliance methodology intended to reduce uncertainty prior to expensive TEMPEST-oriented mitigation deployment.

Unlike conventional TEMPEST evaluation procedures, which are typically performed only after infrastructure deployment or during formal certification stages, the proposed methodology introduces intermediate engineering-level assessment phases derived from publicly available ITU-T recommendations.

The workflow consists of five sequential stages:

1. **Initial Threat Profiling**. Identification of critical assets, expected attacker capabilities, operational frequencies, and applicable intrusion zones according to ITU-T K.84.
2. **Preliminary Spectral Leakage Assessment.** Narrow-band measurements are performed to identify dominant compromising frequencies, harmonics, cable resonances, and propagation-sensitive spectral regions.
3. **Reconstruction-Oriented Evaluation.** Wideband measurements are performed for selected high-risk frequencies and operational modes to evaluate practical information reconstruction feasibility.
4. **Spatial Propagation and Intrusion-Zone Analysis.** Measured emanation levels are incorporated into the proposed propagation model to estimate compromise probability across intrusion zones while considering shielding attenuation, building geometry, and receiver sensitivity.
5. **Targeted Mitigation Optimization.** Shielding reinforcement, zoning adaptation, filtering deployment, and equipment hardening are prioritized according to the measured compromise contribution of individual leakage paths.

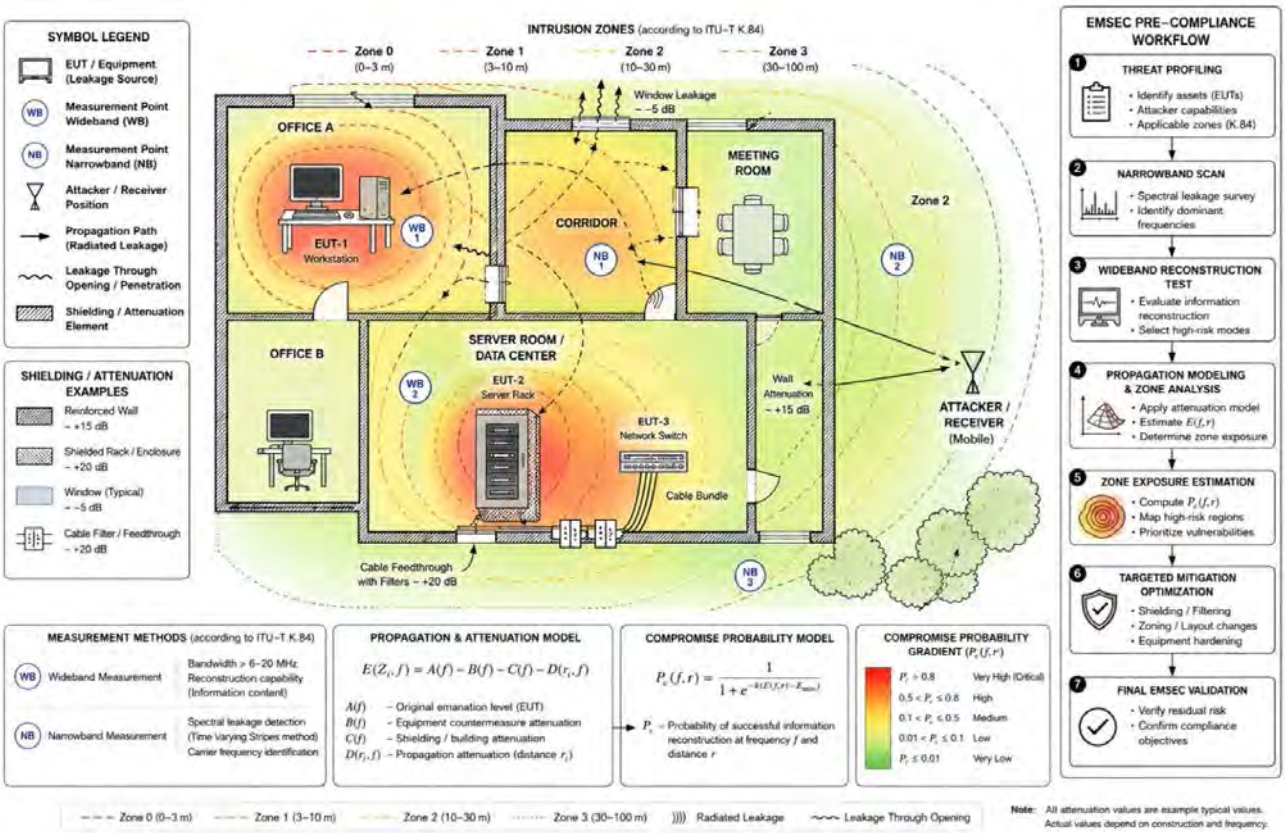


Figure 3. Example facility-oriented EMSEC pre-compliance assessment workflow integrating intrusion-zone analysis, propagation-oriented leakage measurements, shielding evaluation, and targeted mitigation optimization derived from ITU-T K.84 and related recommendations.

The proposed staged approach enables iterative EMSEC hardening where mitigation measures are guided by measurable compromise potential rather than worst-case assumptions alone. Consequently, the methodology supports cost-effective deployment of EMSEC protections while preserving compatibility with formal TEMPEST-oriented evaluation procedures.

4. ANALYTICAL PROPAGATION AND COMPROMISE MODEL

The intrusion zones defined in ITU-T K.84 directly determine the required measurement configuration. As the intrusion distance increases from Zone 0 toward Zone 3, the received field strength decreases due to propagation attenuation and shielding effects.

4.1 Spatial propagation model

For each intrusion zone Z_i, the measurable threat level can be expressed as:

$$E(Z_i, f) = A(f) - B(f) - C(f) - D(r_i, f) \quad (2)$$

where: $E(Z_i, f)$ is the received electric field strength at intrusion zone Z_i and frequency f, $A(f)$ denotes the original compromising emanation spectrum generated by the EUT, $B(f)$ denotes equipment countermeasure attenuation, $C(f)$ represents shielding attenuation, $D(r_i, f)$ represents propagation attenuation associated with attacker distance r_i.

The security criterion becomes:

$$E(Z_i, f) \leq E_{th}(f) \quad (3)$$

which establishes $E_{th}(f)$ the maximum acceptable emanation level for each intrusion zone - the minimum detectable electromagnetic field required for information reconstruction.

Substituting the propagation model yields:

$$A(f) - B(f) - C(f) - D(r_i, f) \leq E_{th}(f) \quad (4)$$

which can be rearranged into:

$$B(f) + C(f) + D(r_i, f) \geq A(f) - E_{th}(f) \quad (5)$$

This expression defines the total required attenuation budget necessary to preserve confidentiality at intrusion zone Z_i. The maximum compromise distance can be derived from the security condition. Solving for distance yields:

$$r_{max}(f) = 10^{\frac{A(f)-B(f)-C(f)-E_{th}(f)}{20}} \quad (6)$$

The obtained value represents the maximum intrusion radius at which compromising information remains detectable.

4.2 Probabilistic Reconstruction Model

Because information reconstruction quality improves gradually rather than instantaneously, EMSEC compromise may be represented using a probabilistic model. The probability of successful information reconstruction is approximated as:

$$P_c(f, r) = \frac{1}{1+e^{-k(E(f,r)-E_{min})}} \quad (7)$$

where: P_c denotes compromise probability, k controls transition sharpness, $E(f, r)$ represents received field strength, E_{min} represents the minimum reconstruction threshold.

This formulation enables continuous estimation of EMSEC exposure rather than binary secure/insecure classification. The wideband and narrow-band measurement methods defined in ITU-T K.84 may now be interpreted within the proposed propagation framework.

The wideband method evaluates practical information reconstruction capability, temporal signal integrity and real compromise feasibility. Conversely, the narrow-band method evaluates spectral leakage localization, carrier frequency identification and frequency-selective propagation characteristics.

5. ILLUSTRATIVE APPICATIONS

5.1 Protection Optimization Concept

The conceptual model shown in Figure 4 illustrates that mitigation strategies implemented without prior leakage characterization may result in excessive shielding complexity and unnecessary infrastructure modifications. In contrast, measurement-driven assessment enables prioritization of dominant leakage paths and frequency components, thereby reducing overall protection cost while maintaining acceptable EMSEC risk levels. The residual compromise probability illustrated in Figure 4 conceptually corresponds to the probabilistic compromise function defined in Equation (7), where mitigation measures progressively reduce the received field strength below the reconstruction threshold. The “Potential Over-Engineering” region illustrates that EMSEC countermeasures implemented without prior leakage characterization may produce disproportionately high shielding and infrastructure costs while yielding only marginal reductions in residual compromise probability. This directly emphasizes the necessity of measurement-driven assessment, since realistic propagation analysis and intrusion-zone evaluation enable optimization of protection measures according to actual electromagnetic exposure rather than worst-case assumptions.

5.2 Spatial-Frequency Exposure Surface

The probabilistic compromise model introduced in Equation (7) can be further extended into a spatial-frequency exposure metric representing the residual EMSEC risk over the operational bandwidth of the compromising emanation. Instead of evaluating information reconstruction feasibility at a single frequency component, practical EMSEC compromise depends on the cumulative contribution of multiple spectral components propagating through the environment.

Accordingly, the average residual compromise probability at distance r may be expressed as:

$$P_{avg}(r) = \frac{1}{BW}\int_{f_1}^{f_2} P_c(f, r)df \quad (8)$$

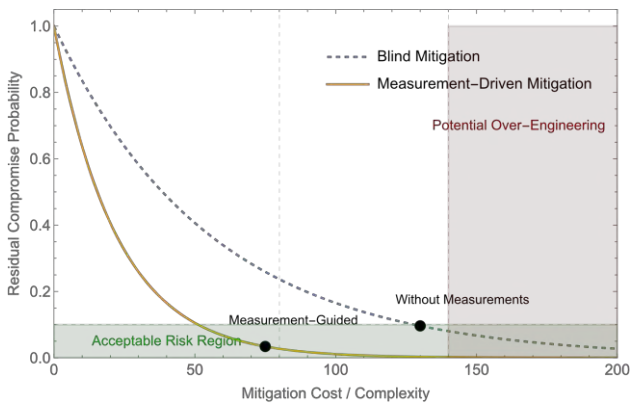


Figure 4. Conceptual comparison between blind EMSEC mitigation and measurement-driven protection optimization.

where: $BW = f_2 - f_1$ denotes the observed frequency bandwidth,

$P_c(f, r)$ represents the probabilistic compromise function,

and $P_{avg}(r)$ expresses the cumulative reconstruction feasibility across the measured spectral interval.

This formulation enables interpretation of EMSEC exposure as a continuous spatial-frequency risk distribution rather than a discrete threshold phenomenon. Consequently, the compromise probability becomes dependent not only on distance and shielding attenuation, but also on the spectral distribution of the emanation source and receiver sensitivity over the operational bandwidth. The proposed model is particularly suitable for visualization through three-dimensional propagation surfaces and intrusion-zone probability maps, Figure 5.
Such visualization enables measurement-driven optimization of shielding deployment, intrusion-zone geometry, receiver-oriented threat assessment and pre-compliance EMSEC mitigation strategies.
The probabilistic compromise surface may additionally be interpreted through intrusion-zone thresholds derived from ITU-T K.84. Instead of representing intrusion zones solely as fixed physical distances, the proposed framework associates each zone with a compromise probability interval over the spatial-frequency propagation surface, Figure 6. Instead of defining intrusion zones only through physical separation distances, the proposed framework associates each zone with a characteristic compromise probability interval derived from the propagation model.

Accordingly, intrusion zones may be represented as:

$$Z_i \equiv \{(f, r) | P_c(f, r) \in \Omega_i\}$$

where: Ω_i denotes the compromise probability interval associated with intrusion zone Z_i and $P_c(f, r)$ represents the probability of successful information reconstruction at frequency f and distance r.

An example interpretation may be expressed as:

– Zone 0: high-confidence reconstruction region ($P_c > 0.8$),

– Zone 1: partial reconstruction region ($0.5 < P_c \leq 0.8$),

– Zone 2: detectable spectral leakage region ($0.1 < P_c \leq 0.5$),

– Zone 3: residual detectability region ($P_c \leq 0.1$).

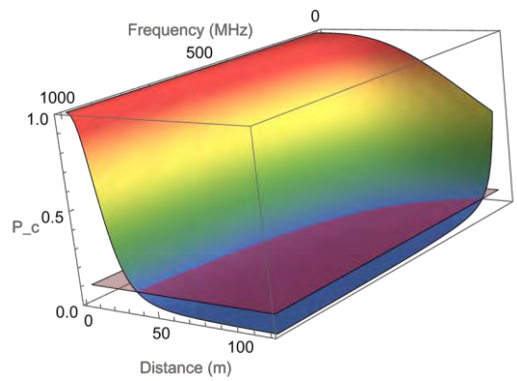


Figure 5. Three-dimensional spatial-frequency representation of probabilistic EMSEC exposure derived from the proposed propagation and reconstruction model. The red plane represents an acceptable residual compromise probability threshold.

This interpretation transforms intrusion zones into measurable propagation-dependent security regions whose boundaries vary according to SE, receiver sensitivity, propagation attenuation and emanation spectral distribution. Consequently, intrusion-zone boundaries are no longer fixed architectural distances, but dynamically depend on the electromagnetic characteristics of the protected environment and attacker capabilities.

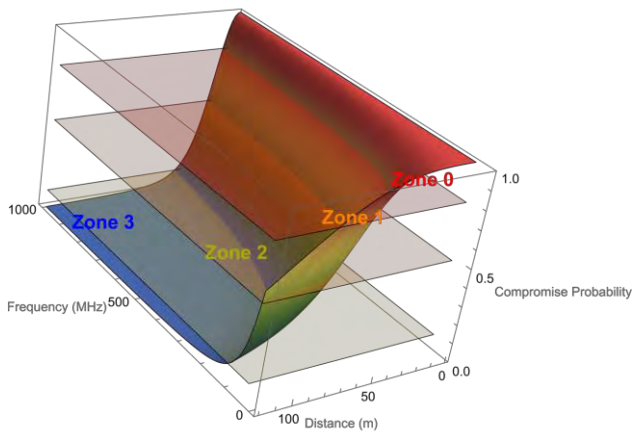


Figure 6. Spatial-frequency EMSEC Exposure with Intrusion Zone thresholds derived from the proposed compromise model.

This representation transforms the intrusion-zone concept into a dynamic propagation-oriented EMSEC assessment model suitable for measurement-driven pre-compliance analysis.

5.3 Intrusion-Zone Exposure versus Shielding

Figure 7 illustrates the relationship between intrusion distance, SE, and the probability of successful electromagnetic information reconstruction. The model assumes an emanation level of 62.2 dBμV/m at the reference distance and a reconstruction threshold of 15 dBμV/m representing the minimum field strength required by the adversary's receiving system. The probability curves are generated using the proposed logistic compromise model, where the parameter k=0.166 controls the transition

between successful and unsuccessful reconstruction regions.

The results demonstrate that increasing shielding attenuation shifts the compromise-probability curves downward and compresses the effective interception range. Without shielding, the probability of successful reconstruction remains relatively high even within extended intrusion zones, whereas shielding levels of 20–30 dB substantially reduce the compromise probability before the Zone 2 and Zone 3 boundaries are reached. This observation supports the central premise of measurement-driven EMSEC assessment: the required protection level is strongly dependent on the actual emanation characteristics and intrusion-zone geometry rather than on generic worst-case assumptions.

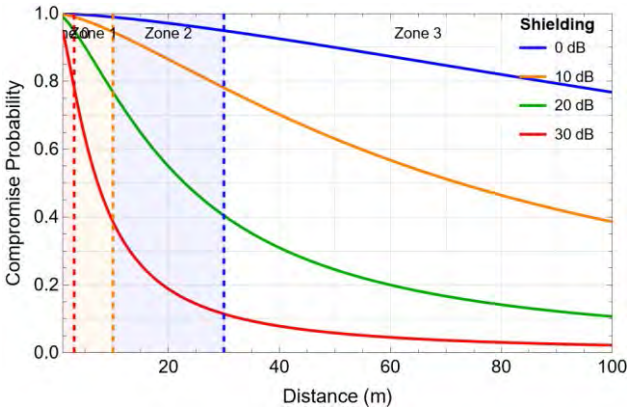


Figure 7. Intrusion-Zone EMSEC Exposure versus Shielding

The selected emanation level (62.2 dBµV/m) approximately corresponds to a leakage source that already satisfies conventional EMC emission requirements, illustrating that EMC compliance alone does not necessarily imply electromagnetic security. The reconstruction threshold (15 dBµV/m) represents the minimum electromagnetic field strength required for meaningful signal recovery by the assumed receiver class, hile the probability sharpness parameter (k = 0.166) models the gradual degradation of reconstruction quality observed in practical EMSEC scenarios. Rather than producing a strict secure/insecure boundary, the model reflects the realistic transition from reliable reconstruction to partial recovery and finally to residual detectability.

The figure highlights that intrusion-zone boundaries alone are insufficient for EMSEC assessment; identical facilities may exhibit substantially different exposure profiles depending on the measured emanation level and SE, thereby justifying the need for pre-compliance EMSEC measurements prior to implementation of costly countermeasures.

5.4 Frequency-Selective Leakage Characterization

In contrast to the spatial propagation analysis presented previously, narrow-band EMSEC measurements focus on identification of spectral components that carry compromising information. As described in ITU-T K.84, specific frequencies may exhibit significantly higher leakage levels due to deterministic electronic processes occurring inside the equipment under test.

The conceptual spectrum shown in Figure 8 models three representative leakage mechanisms frequently encountered in practical systems: clock harmonic, cable resonance and switching noise. Each mechanism contributes differently to the overall compromising emanation profile and therefore requires different mitigation strategies.

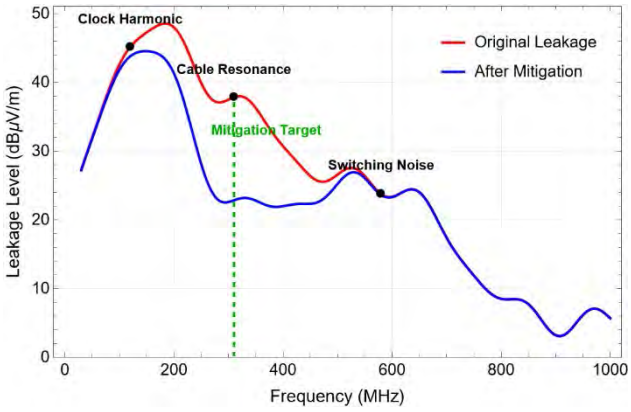


Figure 8. Frequency-Selective EMSEC Leakage Analysis

The resulting leakage spectrum may be approximated as:

$$S(f) = S_{clk}(f) + S_{cab}(f) + S_{sw}(f) \tag{9}$$

where, $S_{clk}(f)$ represents leakage generated by internal clock harmonics, $S_{cab}(f)$ represents resonant amplification caused by connected cables or conductive structures, and $S_{sw}(f)$ represents broadband emissions produced by switching power electronics and digital circuitry. A mitigation function $M(t)$ is then applied around a selected dominant leakage frequency, producing a residual spectrum

$$S_{mit}(f) = S(f) + M(f) \tag{10}$$

which models the effect of shielding, filtering, cable treatment, or redesign actions.

Figure 8 illustrates a representative narrow-band EMSEC measurement scenario derived from the concepts described in ITU-T K.84. The red curve represents the original leakage spectrum measured from an information-processing device, while the blue curve illustrates the residual leakage after application of a targeted mitigation measure. Three dominant leakage components can be observed. The first peak at approximately 120 MHz corresponds to a clock harmonic generated by internal digital timing circuits. The second peak at approximately 310 MHz is associated with cable resonance effects, where connected conductors behave as unintended radiating structures. The third peak at approximately 580 MHz represents switching-noise emissions originating from power conversion circuitry. The dashed line indicates a mitigation target frequency identified through spectral analysis. Application of a frequency-selective countermeasure reduces the leakage level around the dominant resonance, resulting in the mitigated spectrum shown in blue. The figure illustrates how narrow-band measurements can be used to identify dominant compromising frequencies and guide targeted EMSEC mitigation efforts. Instead of applying generalized

shielding measures over the entire spectrum, the proposed methodology uses measurement results to identify dominant leakage frequencies and prioritize mitigation efforts where they produce the greatest reduction in compromise potential. Examples of such mitigation techniques include selective shielding reinforcement, cable rerouting, ferrite suppression, filtering, grounding improvements, or equipment relocation.

Within the context of the proposed analytical framework, the mitigated spectrum corresponds to a reduction of the emanation term $A(f)$ in Equation (5), which directly decreases the maximum compromise distance and lowers the compromise probability defined by Equation (7). The figure therefore provides a practical interpretation of how narrow-band measurements can be translated into measurable reductions of EMSEC exposure and intrusion-zone risk.

6 CONCLUSION

The proposed analytical framework represents a simplified propagation-oriented approximation of practical EMSEC environments. Real electromagnetic leakage scenarios may additionally depend on multipath propagation, polarization mismatch, near-field coupling effects, non-linear leakage mechanisms and receiver-specific reconstruction algorithms. Consequently, the proposed framework should be interpreted primarily as an engineering-oriented pre-compliance assessment methodology intended to support early-stage EMSEC risk evaluation and mitigation prioritization rather than a replacement for formal TEMPEST certification procedures.

To be concluded, the proposed framework demonstrates how publicly available ITU-T recommendations can be combined with propagation-oriented analysis, intrusion-zone assessment, and measurement-driven mitigation planning to establish an engineering-level EMSEC pre-compliance methodology. Unlike traditional TEMPEST evaluation procedures, which are typically performed during certification stages, the proposed approach enables early identification of dominant leakage paths, estimation of compromise probability, and optimization of shielding and zoning strategies prior to expensive infrastructure modifications. However, it is not substitution for formal TEMPEST evaluation. The illustrative examples demonstrate how propagation modeling and spectral leakage characterization can support targeted mitigation while reducing the risk of over-engineered countermeasures.

Acknowledgements

This work resulted from part of research P1/26-28 and was supported by MoD of RS

References

[1] F. LARROCA, P. BERTRAND, F. CARRAU AND V. SEVERI.: *gr-tempest: an open-source GNU Radio implementation of TEMPEST*, 2022 Asian Hardware Oriented Security and Trust Symposium (AsianHOST), Singapore, Singapore, 2022, pp. 1-6, doi: 10.1109/AsianHOST56390.2022.10022149.

[2] C. LAVAUD, R. GERZAGUET, M. GAUTIER, O. BERDER, E. NOGUES, ET AL..: *Whispering devices: A survey on how side-channels lead to compromised information*, Journal Hardware and Systems Security, 2021, 5 (2), pp.143-168. ⟨10.1007/s41635-021-00112-6⟩. ⟨hal-03176249⟩.

[3] Y. LONG, Q. JIANG, C. YAN, T. ALAM, X. JI, W. XU, K. FU: *EM Eye: Characterizing Electromagnetic Side-channel Eavesdropping on Embedded Cameras*, Network and Distributed System Security Symposium, 2024, San Diego, CA, USA

[4] POHL,J., NOACK,A.: *Universal Radio Hacker: A Suite for Wireless Protocol Analysis*, IoTS&P '17: Proceedings of the 2017 Workshop on Internet of Things Security and Privacy, Dallas Texas USA, 2017. https://doi.org/10.1145/3139937.3139951

[5] LIU Z., SAMWEL N., WEISSBART L., ZHAO Z., LAURET D., BATINA L., LARSON M. : *Screen Gleaning: A Screen Reading TEMPEST Attack on Mobile Devices Exploiting an Electromagnetic Side Channel,* Network and Distributed Systems Security (NDSS) Symposium 2021, Virtual, https://dx.doi.org/10.14722/ndss.2021.23021

[6] CAMURATI G., FRANCILLON A.: *Noise-SDR: Arbitrary Modulation of Electromagnetic Noise from Unprivileged Software and Its Impact on Emission Security*, IEEE Symposium on Security and Privacy (SP), 2022, San Francisco, USA, https://10.1109/SP46214.2022.9833767

[7] CAMURATI G., FRANCILLON A. and STANDAERT F.X.: *Understanding Screaming Channels: From a Detailed Analysis to Improved Attacks,* IACR Transactions on Cryptographic Hardware and Embedded Systems (CHES), 2020, https://tches.iacr.org/index.php/TCHES/article/view/8594

[8] ITU-T K.84.: *Test method and guide for EMSEC*, International Telecommunication Union (ITU-T), Recommendation K.84, Geneva, Switzerland, January 2011.

[9] ITU-T K.81.: *High-power electromagnetic immunity guide for telecommunication systems*, International Telecommunication Union (ITU-T), Recommendation K.87, Geneva, Switzerland, August 2024.

[10] ITU-T K.87.: *Guide for the application of electromagnetic security requirements – Overview*, International Telecommunication Union (ITU-T), Recommendation K.87, Geneva, Switzerland, August 2024.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MODEL-BASED DESIGN AND IMPLEMENTATION OF AN OFDM TRANSMITTER FOR STM32F407 VIA MATLAB/SIMULINK

SAŠA GUNDELJ
Military Technical Institute, Belgrade, gundelj@yahoo.com

PAVEL CRNOMARKOVIĆ
Military Technical Institute, Belgrade, crnomarkovic.pavel@gmail.com

SRĐAN TRANDAFILOVIĆ
Military Technical Institute, Belgrade, trandas1@gmail.com

IGOR GORŠIĆ
Military Technical Institute, Belgrade, gorsic@gmail.com

Abstract: *Orthogonal Frequency Division Multiplexing (OFDM) is a widely used multi-carrier modulation technique in modern high-speed wireless and mobile communication systems due to its robustness against multipath fading channels. This paper presents the design, simulation, and hardware implementation of an OFDM transmitter based on Quadrature Amplitude Modulation (QAM) using the STM32F407 microcontroller. The system was fully developed in the MATLAB/Simulink environment, where extensive algorithm validation was performed. Using Simulink Coder, C code was automatically generated and deployed onto the STM32F4DISCOVERY board. This proof-of-concept demonstrates the effectiveness of the model-based design approach for rapid development and validation of communication algorithms on a resource-constrained Cortex-M4-based platform. Spectral measurements of the transmitted signal on a spectrum analyzer showed strong agreement with the simulation results. A receiver has also been implemented, including necessary synchronization blocks; however, due to its complexity, it is not described in detail in this work.*

Keywords: *OFDM, QAM, STM32F407, MATLAB, Simulink, model-based design, embedded communications, proof-of-concept*

1. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) has become the dominant modulation technique in modern high-speed wireless communication systems due to its high spectral efficiency and robustness against multipath fading channels [1], [2]. Traditionally, OFDM transceivers are implemented on high-performance platforms such as Field-Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs), or Software-Defined Radios (SDRs). However, there is a growing demand for deploying communication algorithms on low-cost, resource-constrained embedded microcontrollers for specialized telemetry, Internet of Things (IoT) applications, and educational purposes.

The primary motivation for this work stems from the modernization of legacy telemetry systems operating over the analog FM audio subcarrier of PAL television standard. The objective is to replace traditional Frequency-Shift Keying (FSK) tone-based signaling with a high-throughput 16-Quadrature Amplitude Modulation (16-QAM) OFDM data transmission scheme. To achieve this in a cost-effective manner, the STM32F407 microcontroller (based on the ARM Cortex-M4 architecture), paired with the onboard audio codec of the STM32F4DISCOVERY board, presents an attractive hardware solution. Nevertheless, implementing complex multicarrier modulation on such a platform introduces significant challenges, particularly regarding computational limits, memory constraints, and the critical transition from floating-point simulation to integer arithmetic for hardware interfacing [3].

To address these challenges, this paper adopts a Model-Based Design (MBD) methodology using the MATLAB/Simulink environment. MBD bridges the gap between theoretical algorithm design and embedded deployment by enabling extensive simulation-based verification and automatic generation of optimized C code [4]. This work demonstrates the feasibility of a complete 16-QAM OFDM transmitter operating in real-time on a low-cost microcontroller, detailing the practical MBD workflow from frame-based processing to Inter-IC Sound (I^2S) audio codec interfacing.

2. THEORETICAL BACKGROUND OF OFDM SYSTEMS

The fundamental concept of OFDM lies in dividing a wideband channel into multiple narrowband orthogonal subcarriers, thereby significantly reducing the impact of frequency-selective fading.

Corresponding author: S. Gundelj;
E-mail: gundelj@yahoo.com

DOI: 10.5937/OTEH26083G

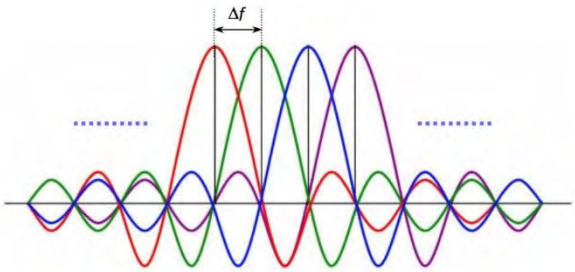


Figure 1. Orthogonal subcarrier spectra in OFDM

The OFDM signal is generated using the Inverse Fast Fourier Transform (IFFT). Let $X[k]$ denote the complex data symbols on the k-th subcarrier ($k = 0, 1, \ldots, N-1$), where N is the total number of subcarriers. The discrete-time OFDM symbol is then expressed as:

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] \cdot e^{j2\pi kn/N},$$

$$n = 0, 1, \ldots, N-1 \qquad (1)$$

After the IFFT operation, a Cyclic Prefix (CP) of length N_{CP} is prepended to the symbol by copying the last N_{CP} samples to the beginning. This mechanism converts linear convolution with the channel impulse response into circular convolution, enabling simple one-tap frequency-domain equalization [5]. In this work, 16-QAM is employed for bit-to-symbol mapping. The 16-QAM constellation allows the transmission of 4 bits per symbol, providing a good agreement between spectral efficiency and noise resilience [6]. The constellation is typically normalized to unit average symbol energy, and Gray coding is applied to minimize bit errors between adjacent symbols.

The main advantages of OFDM over single-carrier modulation schemes include high robustness to multipath propagation, simple frequency-domain equalization, and flexible subcarrier allocation [1]. However, OFDM systems suffer from two notable drawbacks: a high Peak-to-Average Power Ratio (PAPR) and sensitivity to carrier frequency offset and phase noise [7]. These issues require careful system design, particularly when implementing on resource-constrained embedded platforms.

The following sections describe the model-based design of a 16-QAM OFDM transmitter using MATLAB/Simulink, with emphasis on automatic code generation and deployment onto the STM32F407 hardware.

3. MODEL-BASED DESIGN IN MATLAB/SIMULINK

The complete 16-QAM OFDM transmitter was developed using the Model-Based Design methodology within the MATLAB/Simulink environment. This approach enables rapid algorithm development, extensive simulation-based verification, and automatic generation of optimized embedded code [4].

Within the MATLAB/Simulink environment, signal processing can be executed using either sample-based or frame-based processing paradigms. While sample-based processing handles a single scalar value per simulation step, frame-based processing operates on vectors or matrices of data (frames) simultaneously [4]. For the proposed OFDM transmitter, frame-based processing is highly advantageous, as core operations such as the Inverse Fast Fourier Transform (IFFT) and Cyclic Prefix (CP) insertion inherently require block-level data manipulation. Furthermore, adopting a frame-based approach significantly accelerates simulation execution speed and seamlessly maps to the generated embedded C code, effectively mirroring the buffer-oriented data handling utilized by the STM32F407 microcontroller during real-time operation.

Figure 2 presents the complete Simulink model of the 16-QAM OFDM transmitter developed within the MATLAB/Simulink environment. The architecture follows a systematic signal processing chain, beginning with the **Bernoulli Binary Generator** block, which produces a random binary data sequence represented as a column vector of dimension [192×1]. This block serves as the information source for the transmission system.

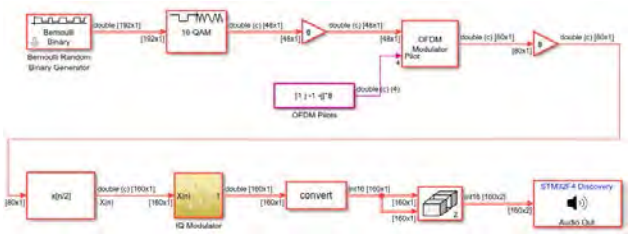


Figure 2. Simulink model of the 16-QAM OFDM transmitter

The binary sequence is subsequently processed by the **Rectangular QAM Modulator Baseband** block, which performs 16-QAM symbol mapping according to Gray coding principles. Each group of 4 bits is mapped to a complex constellation point, resulting in a symbol vector of dimension [48×1]. The 16-QAM constellation is normalized to unit average energy.

Parallel to the main data path, the **OFDM Pilots** block generates known reference symbols of dimension [4×1], which are inserted at predetermined subcarrier positions. These pilot symbols facilitate future channel estimation and frequency offset correction at the receiver side.

The complex symbols are then fed into the **OFDM Modulator** block, which performs the core OFDM processing. This block executes the Inverse Fast Fourier Transform (IFFT) operation, converting frequency-domain symbols into a time-domain signal according to Equation (1). Subsequently, a Cyclic Prefix (CP) is prepended to each OFDM symbol by copying the last N_{CP} samples to the beginning, yielding the final time-discrete OFDM signal:

$$x_{CP}[n] = \begin{cases} x[n + N - N_{CP}], & n = -N_{CP}, \ldots, -1 \\ x[n], & n = 0, 1, \ldots, N-1 \end{cases}, \qquad (2)$$

where N represents the IFFT size and N_{CP} denotes the CP length. The OFDM Modulator block was configured with FFT length N=64, following the subcarrier allocation structure of the IEEE 802.11a/g standard [8], guard bands of 6 lower and 5 upper subcarriers, DC null insertion, and

four pilot subcarriers at indices 7, 21, 43, and 57.

Following OFDM modulation, the signal undergoes amplitude scaling through gain blocks to optimize the dynamic range. The **upsampling block** ($x[n/2]$) then increases the sampling rate by a factor of 2, preparing the signal for digital-to-analog conversion and subsequent IQ modulation.

The processed signal is routed to the **IQ Modulator** subsystem (detailed in Figure 3), where quadrature upconversion to the 8 kHz carrier frequency is performed, with cosine and sine references generated as 160-sample discrete frames at the 32 kHz sampling rate.

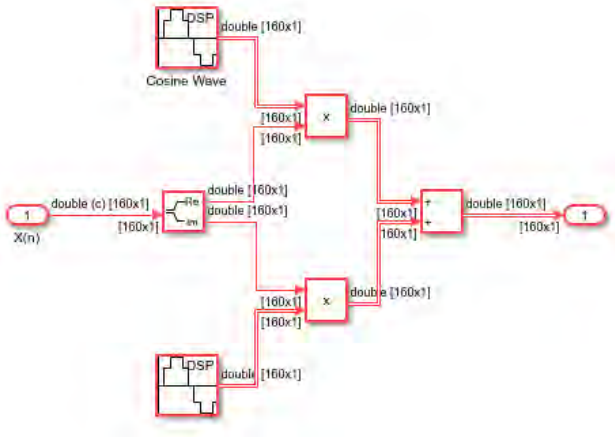


Figure 3. IQ Modulator

Following the IQ modulation stage, the passband signal exists in double-precision floating-point format, which must be converted to 16-bit signed integers to interface with the audio codec connected to the STM32F407. The signal is scaled to utilize the full dynamic range and then quantized to *int16* format, introducing minimal degradation while maintaining sufficient resolution for algorithm validation on the target hardware.

All system parameters, including FFT size (N), CP length (N_{CP}), modulation order ($M = 16$), and sampling frequency (f_s), are defined as tunable workspace variables, enabling rapid exploration of different system configurations during the design phase.

PAPR analysis was not performed in this work, as the primary objective was algorithm validation rather than RF power efficiency optimization.

Channel coding (scrambling, BCH, and convolutional encoding) was omitted from the presented transmitter to prioritize core OFDM signal processing validation. A complementary OFDM receiver has been developed in the same MATLAB/Simulink environment, comprising IQ demodulation, FFT-based OFDM demodulation with pilot extraction, frame and carrier synchronization via PI control loops and pilot-based channel estimation with linear interpolation. Due to space constraints, its detailed description is deferred to future work.

Table 1. System Parameters

Parameter	Value
FFT Size (N)	64
Cyclic Prefix (N_{CP})	16 samples
Guard Bands	6 lower, 5 upper
Pilot subcarriers	4 (indices: 7, 21, 43, 57)
Modulation	16-QAM (Gray code)
Data Subcarriers	48
Sampling Frequency (f_s)	32 kHz
CP Overhead	25%
Interpolation Factor	2
Carrier Frequency (f_c)	8 kHz
Raw Data Rate	38.4 kbps

The specific system parameters used in this implementation are summarized in Table 1. An FFT size of $N = 64$ subcarriers was selected to balance spectral efficiency with the computational constraints of the Cortex-M4 processor. This follows a structure similar to the IEEE 802.11a/g standard [8], where 48 subcarriers are used for data transmission, 4 for pilot symbols, and the remaining are null subcarriers for DC and guard bands. The cyclic prefix length was set to $N_{CP} = 16$ samples (25% overhead), providing sufficient protection against multipath delay spread while maintaining acceptable spectral efficiency. After interpolation the system operates at a sampling frequency of $f_s = 32\ kHz$, which is fully compatible with standard audio codec interfaces. The OFDM modulator itself runs at a base rate of $f_{s,base} = 16\ kHz$. The 16-QAM modulation with Gray coding yields a raw data rate of approximately 38.4 kbps. These parameters were carefully chosen to ensure real-time operation on the STM32F407 while demonstrating the fundamental principles of OFDM transmission.

One of the key benefits of this methodology is the direct path from a verified Simulink model to deployable C code using Simulink Coder and Embedded Coder. After successful simulation, the model was configured for code generation targeting the ARM Cortex-M4 processor, with appropriate optimization settings for execution speed and memory footprint.

4. HARDWARE IMPLEMENTATION AND RESULTS

The generated C code was deployed onto the STM32F4DISCOVERY board featuring the STM32F407VG microcontroller (ARM Cortex-M4, 168 MHz, 192 KB SRAM, 1 MB Flash). The onboard CS43L22 audio codec, interfaced via I²S at 32 kHz, served as the digital-to-analog converter for the passband OFDM signal. No external RF frontend was used; the signal was observed directly at the codec's line output.

Table 2. Memory Footprint

Resource	Used	Available	Percentage
Flash(program)	21.2 KB	1,024 KB	2.1%
SRAM (data)	22.9 KB	192 KB	11.9%

Automatic code generation via Simulink Coder and Embedded Coder produced an executable of approximately 21 KB Flash and 23 KB SRAM footprint, corresponding to 2.1% and 11.9% of the STM32F407VG resources, respectively. The build leverages the ARM Cortex-M4 FPU with hard-float ABI and -O3 optimization, targeting single-precision hardware-accelerated arithmetic. Deployment was performed directly from the MATLAB environment via ST-Link/OpenOCD. Based on the symbol rate of 200 Hz and the observed real-time stability, the CPU utilization is estimated to be well below 5%, leaving substantial headroom for receiver processing.

Figure 4 presents the power spectral density of the transmitted signal obtained from the Simulink simulation, while Figure 5 shows the corresponding measurement captured on the HP 3582A spectrum analyzer at the STM32F4DISCOVERY line output. Both spectra exhibit the expected 64-subcarrier OFDM structure with 250 Hz subcarrier spacing ($\Delta f = f_{s,base} / N = 16000/64$), centered at the 8 kHz carrier frequency. The measured spectrum closely matches the simulation, confirming the fidelity of the implementation and the correctness of the automatic code generation process. Minor deviations in the sidelobe levels are attributed to the finite resolution bandwidth of the analog spectrum analyzer and the reconstruction filter characteristics of the CS43L22 audio codec.

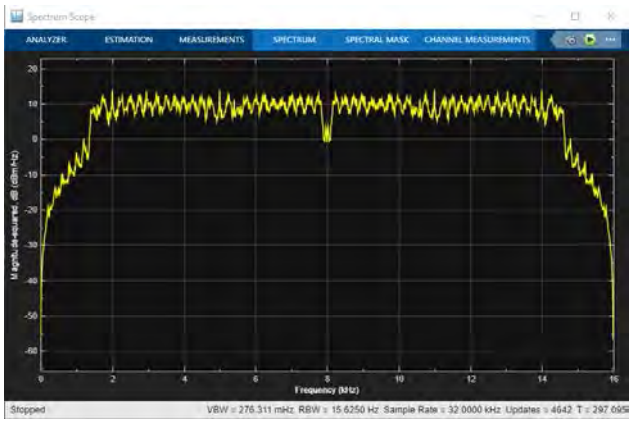


Figure 4. Simulated power spectral density

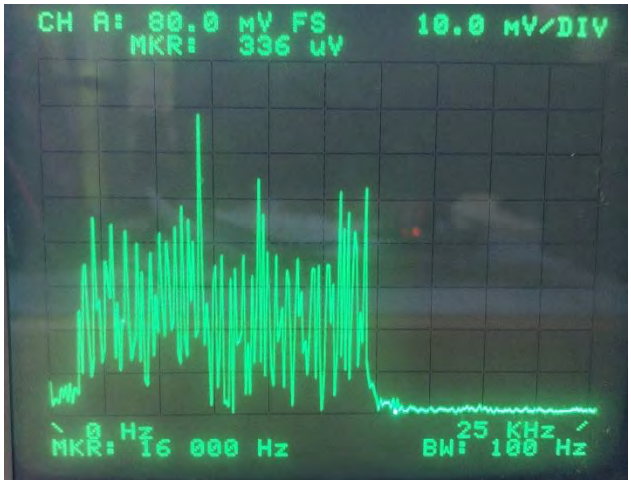


Figure 5. Measured power spectral density

The experimental results validate the model-based design approach: the floating-point implementation preserves the spectral characteristics of the reference simulation, while operating well within the real-time constraints of the Cortex-M4 platform. The primary limitation is the audio codec's 32 kHz sampling rate, which constrains the passband carrier to the audible range — acceptable for wired or near-field telemetry applications, but requiring an external upconverter for true wireless transmission.

5. CONCLUSION

This paper presented a complete model-based design and implementation of a 16-QAM OFDM transmitter on the STM32F407 microcontroller. The system was fully developed and validated in the MATLAB/Simulink environment, with automatic C code generation and deployment onto the STM32F4DISCOVERY board via Simulink Coder and Embedded Coder. Spectral measurements on the HP 3582A spectrum analyzer confirmed strong agreement with simulation results, demonstrating the feasibility of multicarrier modulation on a resource-constrained Cortex-M4 platform. The achieved raw data rate of 38.4 kbps, with minimal CPU and memory utilization, leaves substantial headroom for system expansion. A complementary OFDM receiver with synchronization and channel estimation has also been developed, as has a channel-coded transmitter variant; both are omitted here for brevity. Future work will fully characterize the receiver performance, integrate channel coding, and address fixed-point optimization for enhanced efficiency.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] GOLDSMITH, A.: *Wireless Communications*, Cambridge University Press, Cambridge, UK, 2005.

[2] VAN NEE, R., PRASAD, R.: *OFDM for Wireless Multimedia Communications,* Artech House, Boston, USA, 2000.

[3] ARM LIMITED: *Cortex-M4 Technical Reference Manual*, ARM Limited, Cambridge, UK, 2020.

[4] MATHWORKS: *Model-Based Design for Embedded Systems*, The MathWorks Inc., Natick, MA, USA, 2022.

[5] PROAKIS, J. G., SALEHI, M.: *Digital Communications, 5th ed.*, McGraw-Hill, New York, USA, 2008.

[6] TSE, D., VISWANATH, P.: *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.

[7] BAHAI, A. R. S., SALTZBERG, B. R., ERGEN, M.: *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, Springer, New York, USA, 2004.

[8] IEEE Std 802.11a-1999, Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer in the 5 GHz Band*, New York, USA, 1999.

[9] PASOLINI, G., ZABINI, F., OLIVERI, S.: *A software defined radio platform with Raspberry Pi and Simulink,* European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 2016.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

TELECOMMUNICATION AND INFORMATION SYSTEMS INFORMATION-PSYCHOLOGICAL ACTIVITIES IN MODERN CONFLICTS

JELENA BAJIC
College of vocational studies for Security and criminalistics KIB, Bulevar Cara Konstantina 82-86,NIS,
prof.drjelenabajic@gmail.com

Abstract: *Modern society is founded on both information and disinformation. The Internet and global networking have enabled rapid access to information, but also, in communication terms, introduced significant "noise" within it. Communication is of crucial importance, as those who possess timely information effectively control the situation—a principle that has always held true. Within the military-security operational environment, military communications, radio communications, satellite links, and cyber communication are gaining increasing importance under contemporary conditions, particularly in modern conflicts. In the context of ongoing modern conflicts, the significance of communication is most evident in crisis management and communication during emergency situations. This is especially relevant in defensive military operations, counterterrorism efforts, and responses to hybrid threats, as well as in psychological operations, information warfare, and the corresponding communication strategies employed in military crisis situations. The human factor in defense systems, including the psychological aspects of communication under stress and in military conditions, can significantly contribute to, but also undermine, the effectiveness of operations. All of the above, with particular emphasis on information-psychological operations in modern conflicts, will be discussed in greater detail in the paper.*

Keywords: *information, communication, modern conflicts, psychological operations*

1. INTRODUCTION

The contemporary security environment is characterized by rapid technological transformations that have significantly influenced the planning, execution, and assessment of military and security operations. In addition to the land, maritime, air, space, and cyber domains, the information environment has emerged as one of the key arenas for achieving strategic objectives. Within such an environment, information represents a strategic resource, while the capability to collect, process, protect, and disseminate information has become one of the fundamental elements of modern operations (NATO, 2025).

The development of telecommunication and information systems has enabled near real-time transmission of data, coordination among various levels of command, and decision-making processes based on timely and accurate information. At the same time, digital platforms and social media have created a new information environment in which, alongside reliable information, disinformation, propaganda, and other forms of information influence spread rapidly (NATO, 2023).

A particular challenge stems from the advancement of generative artificial intelligence, which enables the creation of highly convincing textual, audio, and video content. These technologies possess significant potential for improving data analysis processes and supporting decision-making; however, they simultaneously increase the risk of sophisticated disinformation campaigns and manipulation of public opinion (Shoaib et al., 2023).

In this paper, the term *information-psychological activities* is employed as a broader concept encompassing information operations, psychological operations (PSYOPS), strategic communications, and other activities aimed at shaping the perceptions, attitudes, and behavior of target audiences within the information environment. This approach is consistent with contemporary NATO doctrinal documents emphasizing the integration of information-related activities into operational planning and execution (NATO, 2015).

The subject of this study is the analysis of the role of telecommunication and information systems in information-psychological activities during modern conflicts, with particular emphasis on the application of digital technologies and artificial intelligence. The objective of the paper is to examine the importance of these systems in achieving information superiority and to propose a conceptual model linking technological, informational, and psychological dimensions of contemporary information operations.

2. THEORETICAL FRAMEWORK OF INFORMATION-PSYCHOLOGICAL ACTIVITIES

Modern conflicts are characterized by the extensive application of information and communication technologies and an increasing dependence of military and

Corresponding author: J. Bajic
E-mail: prof.drjelenabajic@gmail.com

DOI: 10.5937/OTEH26084B

security systems on the reliable functioning of telecommunication infrastructure. Within contemporary NATO doctrine, information operations are defined as a planning function integrating various information-related activities in order to achieve effects within the information environment (Joint Chiefs of Staff, 2022; NATO, 2015).

In this study, information-psychological activities are conceptualized as an interdisciplinary framework encompassing information operations, psychological operations (PSYOPS), strategic communications, counter-disinformation activities, and other forms of communication-based influence directed toward target audiences. Such an approach facilitates a comprehensive understanding of the technological, informational, and psychological dimensions of modern conflicts.

The foundation of information-psychological activities lies in information superiority, understood as the ability to collect, process, protect, and disseminate relevant information in a timely and effective manner. Information superiority depends not only on the technological sophistication of systems but also on the quality of data analysis, the speed of decision-making processes, and the efficiency of communication among different command levels (NATO, 2015).

Telecommunication systems play a particularly significant role by enabling reliable data transmission among participants involved in operations. Satellite communications, radio systems, optical networks, mobile communications, and digital platforms constitute the basis of modern information activities. Their security and resilience against electronic interference and cyberattacks have a direct impact on the effectiveness of military and security operations (NATO, 2020).

The emergence of social media and digital platforms has further transformed the nature of the information environment. Information now spreads almost instantaneously, while every user can simultaneously act as both a consumer and producer of content. This environment facilitates coordinated disinformation campaigns, propaganda efforts, and other forms of information influence capable of generating substantial psychological and societal consequences (Singer & Brooking, 2018).

An additional challenge arises from the application of generative artificial intelligence, which enables the automated production of highly persuasive textual, audio, and video content. While these technologies offer numerous advantages in data analysis and decision support, they also increase the risk of misuse through the creation of deepfake materials and sophisticated disinformation campaigns (Shoaib et al., 2023).

Contemporary literature suggests that information-psychological activities constitute an interdisciplinary field where telecommunications, information technologies, psychology, communication studies, and security sciences intersect. Understanding the interrelationships among these domains provides the foundation for developing effective mechanisms for protecting the information environment and enhancing societal resilience against modern information threats.

3. INTEGRATED MODEL OF INFORMATION-PSYCHOLOGICAL ACTIVITIES

Contemporary developments in information and communication technologies indicate that the effectiveness of information-psychological activities no longer depends solely on the technical capabilities of systems but increasingly on their integration with information processes and psychological mechanisms influencing perception and decision-making.

An analysis of recent literature demonstrates that these three dimensions are frequently examined separately, whereas their interconnections remain considerably underexplored (NATO, 2015; Joint Chiefs of Staff, 2022). Accordingly, this paper proposes a conceptual model of information-psychological activities integrating three interrelated components: the technological, informational, and psychological dimensions.

The technological component comprises telecommunication and information systems that enable the collection, processing, protection, and dissemination of data. This category includes communication networks, satellite systems, computer networks, command-and-control systems (C4ISR), and artificial intelligence-based platforms. Their primary function is to provide a reliable information environment that serves as the foundation for planning and implementing information-related activities.

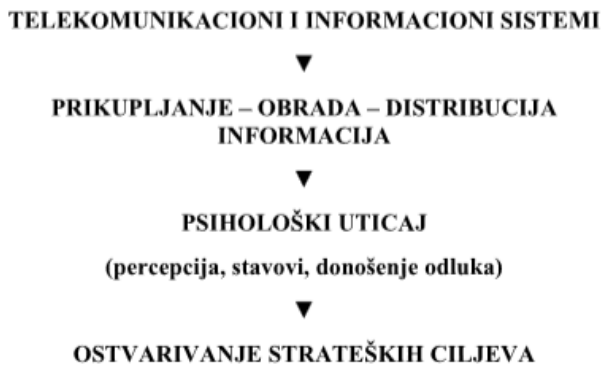
The informational component encompasses the collection, analysis, processing, and distribution of information, as well as strategic communications, information operations, and information content management activities. The purpose of this component is to ensure timely, reliable, and coordinated information supporting the achievement of operational and strategic objectives (NATO, 2023).

The psychological component focuses on understanding how individuals and social groups perceive, interpret, and accept information. Emotions, motivation, previous experiences, trust in information sources, and cognitive biases constitute important factors affecting the effectiveness of information-psychological activities (Singer & Brooking, 2018).

Artificial intelligence plays a central role in connecting all three components. Its application facilitates the processing of large volumes of data, identification of behavioral patterns, assessment of information risks, and support for decision-making processes. At the same time, generative artificial intelligence may be exploited for the automated production of disinformation, deepfake content, and other

forms of information manipulation, further emphasizing the necessity of developing effective mechanisms for protecting the information environment (Shoaib et al., 2023).

The proposed model indicates that the full potential of information-psychological activities can only be achieved through the coordinated interaction of all three components. In this regard, the model may serve as a conceptual framework for future research and the development of modern national security systems.



Slika 1. Integrisani model informaciono-psihološkog delovanja
(izvor: autor)

4. TELECOMMUNICATION AND INFORMATION SYSTEMS IN INFORMATION-PSYCHOLOGICAL ACTIVITIES

Telecommunication and information systems constitute the foundation for planning, implementing, and coordinating information-psychological activities in modern conflicts. Their role extends beyond the traditional transmission of data, as they enable the collection, processing, protection, and dissemination of information that forms the basis for timely decision-making and the achievement of information superiority (NATO, 2015).

The development of digital communication technologies has enabled near-instantaneous information exchange among different levels of command and control. Satellite communications, radio communication systems, optical networks, mobile communications, and Internet infrastructure currently represent the technological backbone of contemporary military and security systems. Their reliability, availability, and resilience against electronic interference and cyberattacks constitute essential prerequisites for the successful conduct of information operations (NATO, 2020).

Particular importance is attributed to integrated Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) systems. By integrating data from multiple sources, these systems facilitate a more accurate understanding of the operational environment and enable more efficient management of available resources. The integration of sensor systems, unmanned platforms, satellite reconnaissance capabilities, and information networks significantly enhances the quality of decision-making under complex operational conditions.

Contemporary information systems increasingly employ artificial intelligence algorithms and big data analytics to identify patterns, assess risks, and support decision-making processes. Their application enables faster information processing, more effective operational planning, and improved forecasting of potential scenarios. However, the same technologies can also be exploited for the automated creation and dissemination of disinformation, thereby further complicating the contemporary information environment (Singer & Brooking, 2018).

In addition to technological advancement, the protection of communication and information infrastructure is of paramount importance. Cyberattacks directed against communication networks, databases, or command-and-control systems may seriously disrupt the functioning of defense and national security institutions. Consequently, the implementation of encryption mechanisms, multifactor authentication, continuous network monitoring, and other cybersecurity measures has become an integral component of modern telecommunication system development (Libicki, 2009).

Based on the analysis presented, it can be concluded that telecommunication and information systems represent a crucial pillar of information-psychological activities, as they enable the attainment of information superiority and significantly influence the planning and execution of activities in modern conflicts.

5. INFORMATION-PSYCHOLOGICAL ACTIVITIES IN MODERN CONFLICTS

Modern conflicts increasingly depend less on the application of conventional military force and more on the ability to manage information and influence the perceptions of target audiences. Information-psychological activities have become an integral element of contemporary military and security practices, utilizing various communication, informational, and psychological instruments to achieve political and strategic objectives (Joint Chiefs of Staff, 2022).

One of the most significant forms of information-psychological activities is represented by strategic communications and psychological operations (PSYOPS), aimed at shaping the attitudes, emotions, and behavior of diverse target groups. Their effectiveness depends on the quality of collected information, knowledge of the psychological characteristics of target populations, and the selection of appropriate communication channels (NATO, 2023).

A major challenge within the contemporary information environment is posed by organized disinformation campaigns conducted through digital media and social networking platforms. Due to the rapid dissemination of content and the possibility of algorithmically targeting specific user groups, disinformation may substantially influence public opinion formation, undermine trust in institutions, and affect political and security-related decision-making processes (Singer & Brooking, 2018).

The development of generative artificial intelligence has further enhanced the capabilities of information-psychological activities. The automated generation of texts, images, audio recordings, and video materials, including deepfake technology, makes it possible to create content that is difficult to distinguish from authentic sources. As a result, assessing the credibility and authenticity of information has become one of the greatest challenges facing contemporary societies (Shoaib et al., 2023).

From a psychological perspective, the effectiveness of information influence largely depends on how individuals perceive and process information. Emotions, previous experiences, cognitive biases, trust in information sources, and perceptions of threat all influence the acceptance or rejection of particular messages. For this reason, contemporary information activities increasingly combine technological capabilities with insights derived from psychology and communication sciences in order to maximize their impact on targeted populations.

It may therefore be concluded that information-psychological activities constitute one of the defining features of modern conflicts and that their importance will continue to increase alongside the further advancement of digital technologies and artificial intelligence.

6. CHALLENGES AND FUTURE DEVELOPMENT PERSPECTIVES

The development of telecommunication and information technologies, digital platforms, and artificial intelligence has significantly expanded the capabilities of information-psychological activities while simultaneously creating new security challenges. Particular threats arise from coordinated disinformation campaigns, the misuse of social media platforms, deepfake content, and sophisticated cyberattacks directed against communication and information infrastructures.

In addition to technological considerations, the human factor remains highly significant. Limited media literacy, insufficient information literacy, and pronounced cognitive biases increase susceptibility to manipulative content. Consequently, strengthening societal resilience requires simultaneous improvements in technological capacities, cybersecurity measures, and continuous user education.

An effective response to contemporary information threats necessitates an integrated approach that combines telecommunication and information systems, strategic communications, psychological resilience, and international cooperation. Under conditions of accelerated artificial intelligence development, it is reasonable to expect that information-psychological activities will assume an increasingly prominent role in the planning and implementation of operations in modern conflicts.

CONCLUSION

Modern conflicts demonstrate that information has become one of the most valuable strategic resources, while telecommunication and information systems play a central role in its collection, processing, protection, and dissemination. At the same time, information-psychological activities are emerging as an indispensable component of contemporary military and security practices, aimed at shaping perceptions, attitudes, and decision-making processes.

The analysis presented in this paper indicates that the effectiveness of information-psychological activities depends on the interaction among technological, informational, and psychological factors. The proposed conceptual model highlights the necessity of considering these dimensions in an integrated manner and may provide a useful framework for future research and the enhancement of information operations.

The continuing development of digital communications, cyber technologies, and artificial intelligence will further increase the significance of information-psychological activities in modern conflicts. Therefore, it is essential to continuously improve telecommunication and information systems, strengthen cybersecurity capabilities, and promote media and information literacy as critical elements of national security.

References

[1] Castells, M. (2010). *The Rise of the Network Society. Wiley-Blackwell.*

[2] Joint Chiefs of Staff. (2022). *Joint Publication 3-13: Information Operations. Washington, DC.*

[3] Libicki, M. C. (2009). *Cyberdeterrence and Cyberwar. RAND Corporation.*

[4] NATO Standardization Office. (2015). *Allied Joint Doctrine for Information Operations (AJP-3.10), Edition A, Version 1.*

[5] NATO. (2020). *Allied Joint Doctrine for Cyberspace Operations (AJP-3.20).*

[6] NATO. (2023). *Allied Joint Doctrine for Strategic Communications (AJP-10).*

[7] NATO. (2025). *Allied Joint Doctrine for the Conduct of Operations (AJP-3).*

[8] NATO Strategic Communications Centre of Excellence. (2022). *A Capability Definition and*

Assessment Framework for Countering Disinformation, Information Influence and Foreign Interference.

[9] Nye, J. S. (2019). *Protecting Democracy in an Era of Cyber Information War. Harvard Kennedy School.*

[10] Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). *Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI and Large AI Models.*

[11] Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The Weaponization of Social Media. Houghton Mifflin Harcourt.*

[12] U.S. Department of the Army. (2018). *ATP 3-13.1: The Conduct of Information Operations.*

[13] Haig, Z. (2020). *Novel Interpretation of Information Operations in Today's Changed Operational Environment.*

[14] Neculcea, C.-A., & Răpan, F. (2023). *"Information Operations – Comparative Doctrinal Analysis." Strategic Impact.*

[15] NATO Cooperative Cyber Defence Centre of Excellence. (2023). *Cyber Defence Reports.*

[16] RAND Corporation. (2022). *Information Operations in Strategic Competition.*

[17] NATO Strategic Communications Centre of Excellence. (2016). *Mapping of StratCom Practices in NATO Countries.*

[18] Joint Chiefs of Staff. (2023). *Joint Concept for Competing.*

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

UAV DETECTION AND TRACKING IN VIDEO USING DEEP LEARNING

LAZAR VELINOV
Iritel a.d, Belgrade, lazar.velinov@iritel.com

JOVAN RADIVOJEVIĆ
Iritel a.d, Belgrade, jovan.radivojevic@iritel.com

Abstract: *Unmanned aerial vehicles (UAVs) pose a growing threat to airspace security, critical infrastructure protection and privacy enforcement. Reliable identification, detection and tracking of small, fast-paced drones in various visual scenes remains an open problem, particularly under varying lighting conditions and sensing modalities. This paper investigates a multi-UAV detection and tracking pipeline from video in special conditions, using a combination of single-stage anchor-free detection with multi-object tracking based on Kalman prediction and hierarchical data association. The proposed pipeline is designed with real-time applicability in mind, balancing inference speed with detection and tracking quality. One of the tasks is to leverage both low-confidence and high-confidence detections during the association stage in order to maintain track continuity under low-contrast conditions and occlusion. The system is evaluated on established benchmarks for UAVs with provided thermal infrared sequences in dense multi-UAV scenarios where UAVs are relatively small in comparison to frame size. Performance assessment is done using standard multi-object tracking metrics, with comparative analysis between alternative tracking strategies. Furthermore, we explore potential improvements on the detection phase targeting small-target recall. The paper discusses the trade-offs between detection accuracy, tracking robustness and computational complexity. The paper is intended for Telecommunications and Information Systems topic.*

Keywords: *UAV detection, multi-object tracking, deep learning, real-time.*

1. INTRODUCTION

Unmanned aerial vehicles (UAVs) over time have become cheaper, more accessible and easier to fly, which has turned them in one of the primary threats to airspace security, both in military and civil environment. Detection and tracking of UAVs emerged as one of leading security problems. Thermal infrared (IR) cameras are widely used in solving this problem due to their performance in complex scenarios such as operating in night or in poor visibility. However, with IR cameras, there is trade-off with UAVs occupying relatively small number of pixels. For example, in Anti-UAV Track 3 [1], dataset used for this research, targets have average width of around 11 pixels in 640×512 frame. Also, IR images carry very little useful information in terms of object texture. Tracking multiple UAVs (or swarm drones) can simply be described as tracking by detection – each frame detector feeds detections to the tracker where these detections are linked to appropriate trajectories. Due to specific nature of the dataset, there are couple of problems that were addressed in this paper. With objects width of average 11 pixels, and no significant texture information, appearance of the object is uninformative. This issue raises a question which motivates this paper: how to achieve good multiple object tracking (MOT) performance with challenging conditions for detection?

To answer this question, we paired anchor-free two detectors YOLOv8s [2] and YOLOX-S [3] with several trackers. Trackers are implemented with following algorithms:

- DeepSORT [4] – algorithm which is based on appearance embeddings
- ByteTrack [5] – run across a range of confidence threshold values
- ByteTrack with camera motion compensation (CMC) – overhead which calculates camera motion, is added to ByteTrack

Each combination of detectors and trackers is evaluated on the same validation split with the HOTA [6] metrics. Higher Order Tracking Accuracy (HOTA) is used because it can be factored to a detection accuracy (DetA) and an association accuracy (AssA), so it lets us evaluate gains and losses in each stage of the pipeline.

This analysis showed that, for this particular problem, detection accuracy is saturated, as DetA score is around 68 for both YOLOv8s and YOLOX-S detectors. Also, association turned out to be binding constrain in obtaining higher HOTA score. However, what stood out the most is increasing AssA by 16 to 20 points when CMC is applied, which consequently raises HOTA by 10 to 12 points, all while keeping pipeline efficiency at a real-time level.

2. METHOD

In this paper we implement tracking by detection paradigm. In each frame independently detector localizes UAVs and tracker links those detections to trajectories while maintaining identity of each UAV. Due to the

Corresponding author: L. Velinov;
E-mail: lazar.velinov@iritel.com

DOI: 10.5937/OTEH26085V

nature of dataset (small objects), we choose anchor-free detectors, which are more suitable for this task.

2.1. Detectors

For detection stage we have chosen two comparable iterations of algorithms from YOLO series: YOLOv8-s and YOLOX-S. Both algorithms are chosen due to their anchor-free design, because anchors could yield worse Intersection-over-Union (IoU) scores for small absolute misalignment (couple of pixels), which could suppress recall. The YOLOX is selected as it is the detector used in the original ByteTrack implementation, and YOLOv8 is selected as an official Ultralytics distribution for comparison. Both detectors are initialized from COCO pretrained weights [7] and fine-tuned once on a training split from the dataset as a single class model. Weights are than frozen so that every tracker uses the same detection front-end for inference. At inference, low-scoring detections are needed for ByteTrack second stage association, therefore confidence threshold is held deliberately low.

2.2. Tracking

In this paper, we evaluate three trackers with per-frame detections from detector inference.

ByteTrack uses constant-velocity Kalman filter [8] to model each track. Detections are associated with each predicted track using IoU metric. A distinguish feature of ByteTrack is its two-stage association. In the first stage, high-confidence detections are matched to tracks. In the second stage, low-confidence detections are used to recover objects so that they are not discarded. Because of small objects nature of the dataset, we expect relatively lower detection confidence in some cases, so we sweep single track initialization confidence threshold τ over values {0.3, 0.4, 0.5}. That parameter controls which detections may be associated with a new track (new track threshold set to $\tau + 0.1$).

DeepSORT uses the same model and augments it with appearance embedding of the object, which is used for re-identifying (Re-ID) temporarily lost tracks. Appearance embeddings are included to test whether appearance is relevant at thermal camera with this resolution and small object scale.

2.3. Camera Motion Compensation

Some thermal video sequences from the dataset are captured with a panning, non-stationary camera. For constant-velocity Kalman filter, that poses a problem, as it cannot differentiate true object motion from global camera motion. While camera is moving, between two frames each track's predicted position shifts away relative to the detection causing low IoU score and breaking the association, which inflates identity switches. To tackle this problem, following the idea of global-motion-compensation (GMC) used in BoT-SORT [9] we insert camera-motion-compensation after motion prediction step. Frame-to-frame camera movement is estimated and each predicted box is warped in new frame's coordinates before matching in association stage.

Transformation is calculated using features from the image. First, we need to detect ORB (Oriented FAST and Rotated BRIEF) [10] keypoints in consecutive frames. Detected keypoints are then matched by descriptor distance and ratio, and fitted to partial-affine transform using RANSAC (Random Sample Consensus) [11] to reject outliers. Resulting affine is applied to Kalman state of each track. Full affine transformation is used for the bounding box centre, an isotropic scale is applied to the height, and linear component is applied to the velocities. Aspect ratio is left invariant.

Due to thermal infrared nature of videos in the dataset, default ORB settings provide insufficient results. The frames are low contrast and untextured, so stable points are either not found or clustered around UAVs (objects whose motion need to be excluded when calculating CMC). Three parameter adaptations are made in order to make ORB engage on the static background. First, Contrast Limited Adaptive Histogram Equalization (CLAHE) [12] is applied before feature detection to expose local structure. Secondly, the FAST (Features from Accelerated Segment Test) [13] corner threshold is lowered from 20 to 7 so that sufficient number of keypoints are yielded from muted background. Finally, we mask out detected UAVs with certain level of confidence (confidence threshold ≥ 0.3) before feature extraction, making sure that keypoints are drawn from the static background rather than moving objects. Confidence threshold needs to be chosen carefully, because low threshold could erase important background features, while high threshold could allow moving objects to be extracted as keypoints. When threshold is optimally set, even if some moving objects aren't masked, RANSAC will remove them as outliers. Implementation of CMC with ByteTrack leaves baseline tracker unchanged as it is implemented additively.

3. IMPLEMENTATION

Experimental setup consists of three parts. First is preparing dataset and creating train and validation splits. Second part is training and inference of detectors. Third and final part consists of evaluation for each combination of detectors and trackers.

3.1. Dataset and split

Dataset used for training and validation consists of 200 thermal infrared videos with resolution of 640×512 at 25 frames per second from Anti-UAV Track 3 training set. Official test set from Anti-UAV Track 3 does not provide annotations for full sequences, but only for the first frame, so it can't be used for tracking evaluation. Therefore, we are using 200 annotated sequences to create splits. Sequences are partitioned into 161 training and 39 validation sequences using density-stratified sampling, so that distribution of UAV density per frame (ranging from 0 to more than 40) is preserved across the splits. Targets are labelled with single class – "uav", and mean target area spans around 11×11 pixels.

3.2. Detector training and inference

Both detectors used in this paper are initialized from COCO pretrained weights and fine-tuned on the training split. Training for both detectors was scheduled for 50 epochs, but halted early once mAP@50 plateaued. YOLOv8 stopped after 15 epochs with best scores at epoch 4, and YOLOX stopped after 12 epochs with best scores at epoch 6. Both detectors were trained on native input resolution (640×512), with SGD optimizer and batch size of 16. Early convergence of mAP@50 for both detectors is an indication that detection accuracy saturates on small sized targets on thermal infrared images. Both detectors exhibit comparable accuracy on validation split. The mAP@50 score is approximately 0.95 for YOLOv8 and 0.94 for YOLOX, while the mAP@50-95 is approximately 0.501 for YOLOv8 and 0.511 for YOLOX. At inference, confidence threshold is held deliberately low at 0.1 and non-maximum suppression IoU of 0.45, retaining the low-scoring detections for ByteTrack's second stage association. Detections per sequence are exported and converted in MOT format, and all tracker share the same detection input so that tracking differences can be attributed to association alone. Training and inference are run on single NVIDIA T4 GPU.

3.3. Evaluation

We evaluate tracking performance with HOTA metrics family using TrackEval [14]. Along with HOTA, we report its factorization as detection part score (DetA) and association part score (AssA). We complement these with MOTA (Multi-Object Tracking Accuracy) [15], IDF1 [16], and count of identity switches made (IDSW). HOTA decomposition is important part of our analysis, as it lets us attribute each gain or loss to a specific stage in tracking pipeline, rather than just reporting aggregate score. All metrics are reported based on evaluation on 39 sequences from validation split.

Since we also evaluate real-time applicability, we also report inference speed and frames per second to measure throughput. We measure entire detection and tracking pipeline end to end, reporting detection and tracking speed separately as well as combined. First, we run a warm up of 20 frames in order to discard CUDA initialization and disk I/O. Speed and frame rate are averaged on 500 frames. As sequences in dataset are 25 frames per second, we take that number as real-time threshold. Benchmark, same as inference, is run on single NVIDIA T4 GPU.

4. EXPERIMENTS AND RESULTS

The performance of each detector-tracker combination on the 39-sequence validation split, evaluated using the metrics described in the previous chapter, is reported in the Table 1. When reading by column, table shows asymmetry in DetA and AssA scores. DetA score is nearly constant, corroborating the claim that detection is saturated in small object thermal infrared regime. However, identity metrics, and AssA which depends on identity, vary significantly across configurations.

For each configuration but one, DetA is nearly identical and lands between 65.89 and 69.46. When comparing two detectors at their best, they are almost indistinguishable, with YOLOv8 scoring 69.23 and YOLOX scoring 69.4. With two different detectors reaching same level of accuracy, and the fact that their training lasted only few epochs due to mAP@50 convergence, we can say that DetA saturation is product of small thermal infrared targets rather than detector capacity.

One configuration that stands out is YOLOv8 with $\tau = 0.5$ (track threshold). Raising τ from 0.3 to 0.5 lowers HOTA from 53.72 to 50.63, which is a result of DetA decline from 67.35 to 56.08, while rise of AssA is insignificant. This is the effect of starving sequences with low-confidence detections – tracker fails to assign new track as $\tau + 0.1$ threshold is too high. YOLOX, whose detector is better calibrated and native to ByteTrack, is less sensitive to τ change (HOTA stays around 57 – there is no substantial gain or loss). For all subsequent comparisons τ is fixed to 0.3.

HOTA variance is almost entirely governed by AssA, which ranges from 43.3 (YOLOv8 + ByteTrack) to 65.0 (YOLOX + ByteTrack + CMC). There is also variance in IDF1 and IDSW metrics, which are implicated by AssA and move along with it.

DeepSORT adds Appearance Re-ID – a learned appearance embedding used in association process. However, at this resolution and undefined texture due to thermal infrared images, appearance Re-ID does not improve association. On YOLOv8, DeepSORT and ByteTrack have identical AssA at 43.3, but DeepSORT has more identity switches (4368 compared to ByteTrack's 3389). On YOLOX, DeepSORT is worse both on HOTA (54.68 vs 57.32) and on AssA (44.92 vs 48.57).

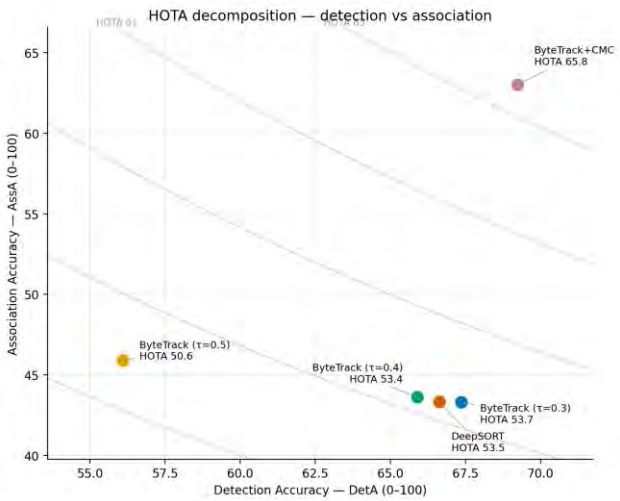


Figure 1. HOTA decomposition for YOLOv8 trackers

Most significant AssA gain is achieved by inserting camera motion compensation before association stage of the pipeline. It raises AssA by 19.68 points on YOLOv8 and 16.44 points on YOLOX, which results in the rise of HOTA by 10-12 points reaching the best results recorded in this experiment (YOLOX + ByteTrack + CMC has

Table 1. Performance of each detector-tracker combination

Detector	Tracker	HOTA	DetA	AssA	MOTA	IDF1	IDSW
YOLOv8	ByteTrack (τ=0.3)	53.72	67.35	43.3	84.92	61.29	3389
YOLOv8	ByteTrack (τ=0.4)	53.35	65.89	43.62	83.19	61.51	3137
YOLOv8	ByteTrack (τ=0.5)	50.63	56.08	45.92	70.63	58.52	2240
YOLOv8	ByteTrack+CMC	65.82	69.23	63.02	87.66	82.05	1134
YOLOv8	DeepSORT	53.49	66.63	43.33	83.16	61.62	4368
YOLOX	ByteTrack (τ=0.3)	57.32	68.22	48.57	85.68	66.97	2620
YOLOX	ByteTrack (τ=0.4)	57.40	68.30	48.66	85.83	67.16	2561
YOLOX	ByteTrack (τ=0.5)	57.02	67.36	48.66	84.74	66.90	2478
YOLOX	ByteTrack+CMC	66.97	69.46	65.01	87.80	83.47	993
YOLOX	DeepSORT	54.68	67.18	44.92	83.89	62.97	3824

HOTA score of 66.97). IDF1 has gains of 21 and 16 points, and identity switches are cut by around two-thirds (3389→1134 on YOLOv8 and 2620→993 on YOLOX). Figure 1 shows HOTA score breakdown in DetA-AssA space for YOLOv8 trackers. Each tracker sits at roughly the same DetA (except ByteTrack with τ = 0.5 which exhibits lower DetA). ByteTrack + CMC advances significantly on AssA axis while maintaining position on DetA axis, placing it in the top-right corner of the plot as the best performing tracker.

Figure 2 shows real-time throughput of each detector-tracker combination. The ones that have the best HOTA score run just above real-time threshold (25 FPS); YOLOv8 + ByteTrack + CMC pipeline runs at around 27 FPS, and YOLOX + ByteTrack + CMC around 28 FPS. CMC adds 15 ms per frame compared to plain ByteTrack, because global motion estimation runs once per frame. DeepSORT throughput collapses to around 3-4 FPS due to adding appearance embedding, which is significantly slower than other trackers, making appearance Re-ID both ineffective and impractical. Most efficient tracker is plain ByteTrack (76.8 FPS on YOLOv8 and 59.6 on YOLOX), but these gains come at the cost of worse association score.

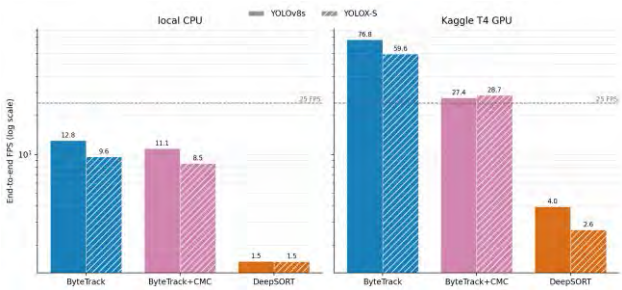


Figure 2. Real-time throughput on different hardware

5. ANALYSIS

In this section, we present key findings and discuss the conclusions drawn. We also explain the rationale behind each intervention and parameter change, as well as the reason behind its success or failure.

Two different detectors used in this research reached similar DetA score (69.23 and 69.46) and converged towards these values after only couple of epochs. Reason for this behaviour of detectors is attributed to the dataset. At box size 11×11 pixels, UAVs are represented with a few warm pixels against cool background. Due to thermal image nature, those bright pixels that represent UAV carry almost no internal structure which would normally be exploited by convolutional layers. Converging in small number of epochs indicate that COCO pretrained weights already scored relatively good, and needed just few epochs to converge. By combining the YOLOv8 detector with ByteTrack and comparing the results with those obtained using YOLOX + ByteTrack (YOLOX is native detector for ByteTrack) we achieved only slightly worse results, which supports the conclusion that detector choice is second-order factor.

This low contrast feature limits not only detection, but also appearance Re-ID feature in DeepSORT. The feature vectors carry little to no useful information. Consequently, consulting appearance cues before motion in the DeepSORT pipeline causes tracker to rely on noisy appearance embeddings, which not only fails to improve association, but also performs slightly worse in terms of AssA and the number of identity switches.

Another consequence of small-sized low-contrast targets is low detection confidence. That directly affects the ByteTrack's initialization of new tracks. Gate for initialization of new tracks is τ + 0.1, which for ByteTrack's default track threshold of τ = 0.5 requires detection confidence to be more than 0.6 to initialize a track. Default threshold was proven to be too high on this dataset, as some sequences contributed almost no true positives, collapsing DetA, MOTA and HOTA on YOLOv8. Lowering the τ to 0.4 and 0.3 is enough to recover scores on those sequences starved by high threshold, while not harming the healthy ones. Better-

calibrated detector YOLOX was proven to be insensitive on changing τ values from 0.3 to 0.5.

Addition of CMC is decisive intervention in this research. The constant-velocity Kalman filter is used to model motion in image coordinates and it doesn't take camera movement into consideration when modelling motion. If camera pans, entire frame of reference is shifted. As a result, camera displacement introduces a discrepancy between the predicted and detected bounding boxes. Consequently, IoU-based association fails, and new identity is assigned to the same object. Adding CMC to detection and tracking pipeline removes that discrepancy by estimating global inter-frame transform. It warps predicted states into the new frame, so that association stage has true object-only motion, not object and camera motion. Since CMC intervention only repairs coordinate frame when warping predicted states, almost all the gain is achieved in association stage (Figure 1).

6. CONCLUSION

In this research, we studied detection and tracking of multiple UAVs from thermal infrared videos in extreme small object regime. We used HOTA decomposition to attribute gains and losses to detection versus association. HOTA decomposition proved that, across combinations of two anchor-free detectors and several trackers, detection saturates regardless of the architecture, while association is the binding constraint. We found that, within small object thermal regime, appearance embeddings (appearance-based re-identification) do not improve over motion-only baseline, while sacrificing performance resulting in approximately 20 × lower throughput. Decisive intervention is adding camera motion compensation to baseline ByteTrack. Global motion estimator applied before association stage raises AssA by up to 20 points, cuts number of identity switches by roughly two-thirds, while maintaining a frame rate above the real-time threshold (YOLOv8 + ByteTrack + CMC run at 27.4 FPS, while video capture rate is 25 FPS), making this system deployable for real-time applications.

The conclusions drawn in this paper should not be considered applicable on UAV detection and tracking in a broader context. Rather, they should be interpreted in light of the specific characteristics of the dataset – thermal infrared low resolution pictures and small objects as targets. The motion model uses partial-affine global transformation. It ignores 3D motion and parallax, and only assumes planar camera motion as dominant. As these sequences mainly consist of distant aerial targets, this motion model is an adequate approximation, although it limits generality of our approach.

References

[1] China Telecom, "Official Dataset Release for the 4th Anti-UAV Challenge". China Telecom, Mar. 03, 2025. doi: 10.5281/zenodo.15103888.

[2] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO (YOLOv8)," version 8.0.0, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[3] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, 'YOLOX: Exceeding YOLO Series in 2021', *arXiv [cs.CV]*. 2021.

[4] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric." Accessed: Jul. 07, 2026. [Online]. Available: https://arxiv.org/abs/1703.07402

[5] Y. Zhang *et al.*, 'ByteTrack: Multi-object Tracking by Associating Every Detection Box', in *Computer Vision -- ECCV 2022*, 2022, pp. 1–21.

[6] J. Luiten *et al.*, "Hota: A higher order metric for evaluating multi-object tracking," *International journal of computer vision*, vol. 129, no. 2, pp. 548–578, 2021.

[7] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, Cham: Springer International Publishing, 2014, pp. 740–755.

[8] R. E. Kalman, 'A New Approach to Linear Filtering and Prediction Problems', *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 03 1960.

[9] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust Associations Multi-Pedestrian Tracking." Accessed: Jul. 07, 2026. [Online]. Available: https://arxiv.org/abs/2206.14651

[10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981, doi: 10.1145/358669.358692.

[12] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems IV*, pp. 474–485, 1994.

[13] E. Rosten and T. Drummond, 'Machine Learning for High-Speed Corner Detection', in *Computer Vision -- ECCV 2006*, 2006, pp. 430–443.

[14] J. Luiten and A. Hoffhues "TrackEval." 2020. [Online]. Available: https://github.com/JonathonLuiten/TrackEval

[15] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," EURASIP Journal on Image and Video Processing, vol. 2008, no. 1, p. 246309, 2008.

[16] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, 'Performance Measures and a Data Set for Multi-target, Multi-camera Tracking', in *Computer Vision -- ECCV 2016 Workshops*, 2016, pp. 17–35.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EXTENDING THE XTTS MODEL FOR SERBIAN TEXT-TO-SPEECH GENERATION

STEFAN IVANOVIĆ
Serbian Armed Forces, Center for Applied Mathematics and Electronics, Belgrade, stefanivanovic1012@gmail.com

DIMITRIJE KOLAŠINAC
Serbian Armed Forces, Center for Applied Mathematics and Electronics, Belgrade, dimitrije.kolasinac@vs.rs

JORDAN ATANASIJEVIĆ
Serbian Armed Forces, Center for Applied Mathematics and Electronics, Belgrade, jordan.atanasijevic@vs.rs

KRISTINA ŽIVANOVIĆ
School of Electrical Engineering, University of Belgrade, Belgrade, zk245024p@student.etf.bg.ac.rs

Abstract: *This paper presents an extended version of the XTTS model, a multilingual open source Text-to-Speech (TTS) model, with Serbian language Text-to-Speech Generation. This is achieved by fine-tuning the original XTTS model using open source CoquiTTS framework with a transcribed speech dataset of Serbian speakers. Apart from that, the tokenizer of the CoquiTTS framework is extended to support the Serbian language. The dataset used for fine-tuning in this paper is created using YouTube videos in the Serbian language with Creative Commons (CC) licenses. YouTube videos are transcribed with the Sagicc/faster-whisper-medium-sr model from Hugging Face, which is a fine-tuned version of the Faster Whisper Medium, Automatic Speech Recognition (ASR) model, for the Serbian language. Transcript generation is performed on a GPU with 6 GB of VRAM. Generated transcripts are checked and corrected manually. The dataset creation process resulted in about 140 hours of Serbian audio-transcribed data. The fine-tuning process is performed using the described dataset on a Kaggle notebook, using a session with a GPU P100 with 16 GB of VRAM, and it took around 25 hours for the whole training process. The final result is open source model capable of generating natural-sounding speech in the Serbian language.*

Keywords: *XTTS model, Text-to-Speech (TTS), Serbian language, dataset, fine-tuning*

1. INTRODUCTION

Text-to-speech (TTS), as an important field in Natural Language Processing (NLP) and generative AI, has witnessed an amazing revolution in the last five years following advancements in the sphere of deep learning. However, despite the presence of some truly wonderful and expressive synthetic voices in high-resource languages that are almost indistinguishable from humans, low-resource languages such as Serbian simply fails to keep that pace. It becomes even more evident when taking into account the fact that there is absolutely no good open-source TTS system available for Serbian. The main goal of this paper is creation of open-source TTS system that will bring closer Serbian TTS to TTS of high-resource languages.

Considering the fact that most modern TTS model's architectures has no corelation with its languages learning capabilities and that model capabilities in certain language solely depend on quality of training data, everything boils down to the fact that the main bottleneck is scarcity of data. Also, existing open source models used for other languages can be used for Serbian language as well. Main missing part of the puzzle for Serbian open-source TTS system is not model, its scarcity of open source Serbian audio datasets. That is secondary goal of this paper, creation of Serbian audio transcribed dataset. The main goal of the paper is achieved using created dataset in fine-tuning process on existing open source TTS model.

2. TTS MODEL

Following expansion of generative AI, researches applied new principles in the field of TTS. Most of the modern TTS models use LLM based architecture with adaptations for working with audio data. In recent years many different approaches and improvements of TTS model architecture emerged [1]. First and important step in building Serbian TTS system is selection of open-source TTS model that we used for fine-tuning training process.

2.1. TTS model selection

Models are searched on Hugging Face platform and following specifications are taken into account: size of the model, resource requirements, speed and quality of generated speech, multilinguality, usage amount, popularity and its documentation for fine-tuning process. XTTS v2 model is picked for this work which is also part

Corresponding author: S. Ivanović
E-mail: stefanivanovic1012@gmail.com
DOI: 10.5937/OTEH26086I

of CoquiTTS framework. Main reason for that is its multilinguality and support for 16 languages [2], which is displayed in Table 1 with data amount used for original XTTS training. This is important feature because in the original training process of the model, model learnt common patterns and rules for all languages, especially similar Slavic languages (Polish, Russian and Czech), which can be reused for Serbian language which leads to smaller dataset required for fine-tuning the model with Serbian data. Other reasons for picking XTTS v2 are its good documentation [3] for fine-tuning process and finaly capabilities to resource requirements ratio.

Table 1. Number of hours for each language in XTTS original training dataset [2]

Language	Hours
English	14513.1
German	3584.4
Spanish	1514.3
French	2215.5
Italian	1296.6
Portuguese	2386.8
Russian	147.1
Dutch	74.1
Czech	52.4
Korean	539.1
Hungarian	62.0
Japanese	57.3
Turkish	165.3
Arabic	240.9
Chinese	233.9
Polish	198.8
Total	27281.6

2.2. XTTS Architecture

The XTTS model is based on the TorToise model, with specific modifications to improve performance in multilingual training and zero-shot TTS (ZS-TTS), as well as to enhance training and speech generation speeds.

During the training process, the XTTS model is fed with text and the corresponding audio recording of that text, serving as the desired output for learning, along with a reference audio recording of the speaker, which allows the model to learn that specific speaker's voice characteristics. Upon entering the model, both audio recordings are converted into Mel spectrograms, which are subsequently processed through the model's blocks. Figure 5 shows a diagram of the XTTS model, illustrating the components of the XTTS architecture itself [2].

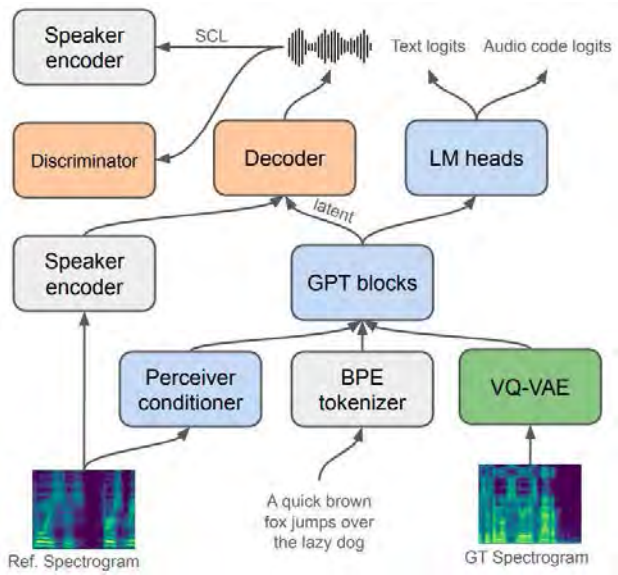


Figure 1. XTTS training architecture [2]

During the training process, the Mel-spectrogram corresponding to the audio associated with the input text is passed through a VQ-VAE (Vector Quantized-Variational AutoEncoder) block. The function of this block is to encode the input Mel-spectrogram into a more compressed form by encoding each spectrogram frame using a codebook containing 8,192 codes. The codebook is filtered to retain only the 1,024 most frequent codes; this represents one of the changes compared to the TorToise model. The authors state that this process has increased the model's expressiveness.

The input text is tokenized using a BPE (Byte Pair Encoding) tokenizer, the same one employed in GPT models. At the core of the model lies a GPT-2 block tasked with generating corresponding VQ-VAE codes from the text tokens. A reference Mel-spectrogram is processed through a conditional encoder block, yielding a vector representation of the speaker's voice characteristics extracted from that spectrogram. This vector representation is also fed into the GPT-2 block, thereby conditioning the generation of VQ-VAE codes on the specific speaker's characteristics. Finally, the model's output stage features a decoder based on the HiFi-GAN vocoder.

The authors state that, instead of VQ-VAE codes, they used the latent space of the GPT-2 model as input to the decoder, similar to the approach taken with the TorToise model. The reason for this is that VQ-VAE codes are highly compressed, and their direct use results in speech generation containing pronunciation errors. Decoder audio generation is also conditioned on the vector representation of a specific speaker's voice characteristics, obtained using the H/ASP model [2].

3. DATASET CREATION

As it is stated at the beggining of the paper, dataset for fine-tuning process should consists of transcribed audio data of Serbian speakers. The amount of that type of data for low resource languages in the open-source community

is even smaller than that of the plain textual data which is also lacking.

3.1. Challenges of Serbian language

Regarding the impact of the Serbian language's complexity on the development of TTS systems, pitch accents pose the greatest challenge among all its characteristics. In Serbian, words with the same written form can carry different meanings depending on their accentuation. From the perspective of a TTS system, there must be a mechanism to convert a word with a specific written form into the appropriate audio output with the correct accent. The system itself requires a mechanism to differentiate words, not merely based on their written form but also on context, in order to distinguish between words based on meaning and generate an accurately accented audio output. This was the reason for picking XTTS model, which is LLM based TTS system which has proven to have the best architecture for data with complex context inter relation.

3.2. Data preparation

In contrast to ASR models, which perform better when trained on noisy audio data, TTS models require audio recordings with minimal noise. This factor complicates the creation of a suitable dataset for this task. Given the scarcity of existing datasets that meet the necessary criteria, a suitable dataset was created specifically for this study. The crated dataset is named SLYIVA-D1 (Serbian Language YouTube Integrated Verbal Audio Dataset v1) and process of its creation is described in the following paragraphs.

Such a dataset can be created by recording audio content and transcribing it, or by collecting data from the Internet under licenses that permit the use of content for training artificial intelligence models without infringing on copyright. Inspired with YODAS (Youtube-Oriented Dataset for Audio and Speech) dataset [4], a second approach was used to create the necessary dataset. When dealing with dataset creation, especialy when data is collected from the Internet in various types and formats, data should be preprocessed in several steps (Figure 2), which are described in details in the following paragraphs.

Content from the YouTube platform bearing a Creative Commons (CC) license served as the source of raw data. Videos in Serbian with a Creative Commons (CC) license were selected [4]. The chosen videos consisted of lectures, speeches, or audiobooks to ensure the required audio quality and speaker voice consistency throughout the recording. The videos were downloaded in .webm format, and the filenames included speaker pseudonyms that were subsequently used to label the speakers for the training audio samples.

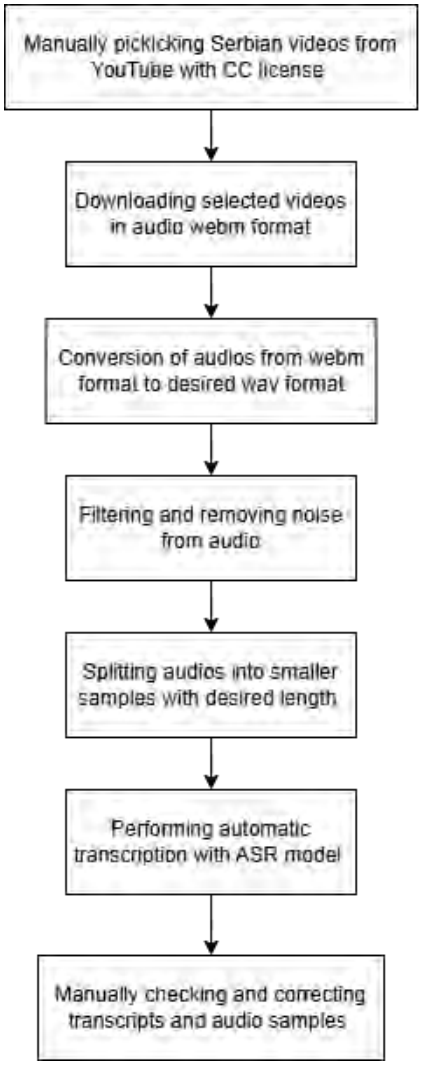


Figure 2. Diagram of steps applied in process of SLYIVA-D1 creation

As specified in the CoquiTTS documentation for the XTTS model, training audio recordings must be in .wav format with a sampling rate of 22,050 Hz, single-channel (mono), and a 16-bit sample depth. The training data format itself matches the format used in the LJSpeech public dataset for English (example in Figure 3 and Table 2). Regarding the characteristics of the data itself, achieving better results requires the data to possess the following attributes:

– Normal distribution of audio recording lengths (3-11s)
– Absence of noise and errors
– Speaker consistency across all audio recordings
– Good phoneme coverage for the specific language
– Naturalness of speech in the audio recordings [3]

```
dataset/
├── metadata.csv
└── wavs/
    ├── audio_0001.wav
    ├── audio_0002.wav
    ├── audio_0003.wav
    └── …
```

Figure 3. Example of LJSpeech dataset format – Folder structure of the dataset

Table 2. Example of LJSpeech dataset format – Content of metadata.csv file

Audio file relative path	Transcript	Speaker name
wavs/audio_0001.wav	Today we will speak about text to speech models	Jovana
wavs/audio_0002.wav	It's purpose is to synthesize speech from text data.	Milan
wavs/audio_0003.wav	Modern models are based on transformer architecture.	Marko

Conversion from .webm format to the required format was performed using the ffmpeg tool. Due to the presence of noise in the audio recordings, stemming from background noise, microphone self-noise, and the quality of the recording itself, noise reduction was performed. The first step in this process involved passing the audio recordings through a high-pass filter to eliminate low-frequency noise falling outside the human voice's frequency spectrum. A fifth-order Butterworth filter with a cutoff frequency of 80 Hz was used for this purpose. The Python library "noisereduction" was used to remove noise from the frequency spectrum overlapping with the speech spectrum. At the very end of this process, Root Mean Square (RMS) normalization was performed to eliminate any amplitudes exceeding 1.0 that might occur after using the "noisereduction" library.

After converting to the desired format and filtering out noise, the next step is to split large audio recordings into smaller segments to meet the duration requirements for the dataset samples (Figure 4). It is important that the segmentation process does not disrupt the speaker's flow of speech, specifically that the original audio is not cut off in the middle of a word. Ideally, the split points should coincide with longer pauses in the speech. Data regarding the timestamps of video segments containing speech were extracted using the Silero Voice Activity Detection (VAD) model via the `silero_vad` Python library. The resulting segments varied in length, since the XTTS fine-tuning process requires audio samples between 3 and 11 seconds long, shorter segments were merged to meet this requirement. Conversely, longer segments were subjected to further processing to shorten them. Such long segments typically occurred when the speaker had a prolonged speaking turn or when the Silero VAD model failed to detect the end of speech. In these instances, an alternative method was employed to further subdivide the segments. The Sagicc/faster-whisper-medium-sr Automatic Speech Recognition (ASR) model was used for transcription, a topic to be discussed further later on. This model was also employed to further segment longer audio clips. In addition to the transcript itself, the model generates corresponding timestamps for each word. These timestamps were used to identify longer pauses and select suitable cut points. This method is less reliable than using a VAD model because the generated timestamps are often offset by several tens or hundreds of milliseconds, which can result in a cut point being selected in the middle of a word. This is also the reason why this method was used as a secondary option. For this reason, the newly segmented parts were specially marked so that particular attention could be paid to them during the manual review and correction process.

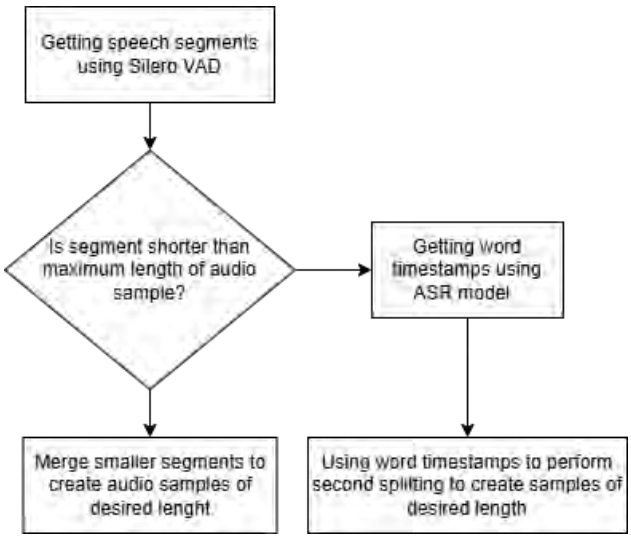


Figure 4. Flowchart of audio splitting process

Finally, once the audio recordings of the desired length were obtained, they needed to be transcribed. It is important to note that there is no open-source Automatic Speech Recognition (ASR) model capable of producing a perfectly accurate transcription of the Serbian language and that such accuracy can only be achieved through human effort. To accelerate the transcription process and minimize the amount of manual work required, a semi-automated approach was adopted. Initial transcription was performed using an ASR model, followed by a final manual review and error correction. For this purpose, the Sagicc/faster-whisper-medium-sr, model from the Hugging Face platform, was selected. This model is a fine-tuned version of Faster Whisper Medium, an ASR model with enhanced capabilities for the Serbian language achieved through fine-tuning [5]. Whisper models were developed by OpenAI and are based on the encoder-decoder Transformer architecture. Since this model architecture supports parallelization, efficient operation requires running it on a machine that supports the CUDA platform [6]. Transcript generation was performed on an NVIDIA GeForce RTX 4050 Laptop GPU with 6 GB of video memory and CUDA installed. The resulting transcript, paired with a pseudo-speaker name and the audio recording, constitutes a single dataset sample. The data was packaged in the expected LJSpeech format, as specified in the XTTS model documentation.

As it is mentioned in previous paragraph Sagicc/faster-whisper-medium-sr model is not perfect, it exhibits a Word Error Rate (WER) of 7.38% during transcription, as noted by its author [5]. Because of that, the audio samples were manually verified and corrected. Normalization of the transcripts was also required. To facilitate this, a simple GUI application was created using the gradio

Python library. The application allows for browsing the dataset samples and making any necessary modifications to them. A semi-automated approach was used for number normalization, the application suggests a normalized text to the user, though errors in grammatical case forms may occur and require user correction (Figure 5). Users also have the option to trim superfluous segments of the audio recording (such as stammers, background voices or sounds) that would otherwise compromise training quality (Figure 6). This marks the final step in the dataset creation process for this task, resulting in an audio dataset (in LJSpeech format) comprising approximately 140 hours of audio, complete with corresponding transcripts and speaker labels, which was subsequently used for fine-tuning.

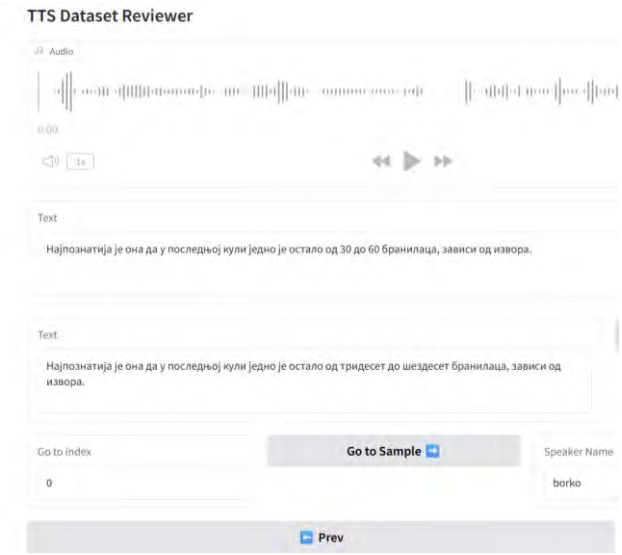


Figure 5. Part of the app screenshot showing its audio playing, transcript editing and number normalization suggestion functionalities

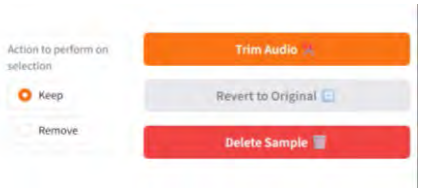


Figure 6. Part of the app screenshot showing its audio crop functionality

4. FINE-TUNING PROCESS

Since the core of the XTTS model is a GPT-2 model based on the Transformer architecture, its efficient use requires running the model on a GPU [2]. Fine-tuning the XTTS model requires a minimum of 16 GB of video memory. For this reason, the fine-tuning process for the Serbian language was carried out in a Kaggle notebook using a session with a P100 GPU featuring 16 GB of video memory. The example for extending the model used in this work was found on GitHub project anhnh2002/XTTSv2-Finetuning-for-New-Languages [7].

The project creator used this example to extend the XTTS model to the Vietnamese language [7]. As the project is no longer actively maintained, modifications were made to the project's scripts for the purposes of this study and to ensure compatibility with newer versions of Python and its libraries. Since the tokenizer rules within the CoquiTTS project are statically defined, the tokenizer itself was modified to support sentence tokenization in Serbian.

Subsequently, a script was executed to expand the XTTS model's vocabulary. Using a tokenizer, this script iterates through the entire text in the dataset and adds new tokens to the vocabulary. Finally, the training script was run with the following parameters:

– Number of epochs – 5
– Number of samples per training step – 4
– Gradient accumulation – 4
– L2 regularization – 0.01
– Learning rate – 0.000005

The fine-tuning process took approximately 25 hours, resulting in an XTTS model capable of generating natural speech in Serbian.

5. CONCLUSION

The primary product of this paper is fine-tuned XTTS capable of syntesizing Serbian speech. This version of model is not evaluated or compared with other Serbian open source or commercial TTS solutions. That can be done using subjective and objective methods which are not well established and standardized for Serbian language. Also every other model has its usage documentation which makes this task more time consuming and complex. This impose some future directions of development and research. Firstly, making standardized objective and subjective evaluation methods for Serbian TTS model, which than can be used to evaluate this as well as other TTS models.

Secondary product is SLYIVA-D1 dataset, which can be used for any other training or fine-tuning process of TTS or ASR models. For future work methodology applied for SLYIVA-D1 can be combined with YODAS methodologies [4] which can be used to create larger variant of this dataset.

Acknowledge

This research was funded by the CAME (Center for Applied Mathematics and Electronics) projects No O1/25-27 and No P1/23-28.

References

[1] YANG, Z., SHIMIZU, S., YU, Y., CHU, C.: *When large language models meet speech. A survey on integration approaches*, Findings of the 63rd Annual Meeting of the Association for Computational Linguistics *(ACL)*, 2025.

[2] CASANOVA E., DAVIS K., GÖLGE E., GÖKNAR G, GULEA I, HART L., ALJAFARI A., MEYER J., MORAIS R., OLAYEMI S., WEBER J.: *XTTS: a*

Massively Multilingual Zero-Shot Text-to-Speech Model, Proceedings of the Interspeech, pp. 4978–4982, 2024.

[3] COQUI TTS CONTRIBUTORS: *Coqui TTS: A library for advanced Text-to-Speech generation* (Version 0.28.0.dev0) [Computer software]. https://coqui-tts.readthedocs.io/en/latest/, 2021.

[4] LI, X., TAKAMICHI, S., SAEKI, T., CHEN, W., SHIOTA, S., WATANABE, S.: *YODAS: Youtube-Oriented Dataset for Audio and Speech*, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Taipei, Taiwan, 2023.

[5] SAGICC: *faster-whisper-medium-sr*, Hugging Face, Repository https://huggingface.co/Sagicc/faster-whisper-medium-sr, 2023.

[6] RADFORD, A., KIM, J. W., XU, T., BROCKMAN, G., MCLEAVEY, C., & SUTSKEVER, I.: *Robust speech recognition via large-scale weak supervision*, International Conference on Machine Learning (pp. 28492-28518). PMLR, 2023.

[7] NGUYEN, A. H.: *XTTSv2 Finetuning for New Languages* https://github.com/anhnh2002/XTTSv2-Finetuning-for-New-Languages [Source code]. GitHub, 2024.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PASSIVE TIME OF ARRIVAL BASED TARGET LOCALIZATION FOR RADAR SYSTEMS USING HYBRID METAHEURISTIC OPTIMIZATION ALGORITHMS

MAJA ROSIĆ
Faculty of Mechanical Engineering, University of Belgrade, Belgrade, mrosic@mas.bg.ac.rs

MILOŠ SEDAK
Faculty of Mechanical Engineering, University of Belgrade, Belgrade, msedak@mas.bg.ac.rs

Abstract: *Accurate target localization of passive target remains a fundamental challenge in modern telecommunications and defense systems. Passive target localization based on time of arrival (TOA) measurements represents a critical capability in modern radar and military surveillance systems, enabling accurate position estimation without active signal emission by the target. Under line of sight (LOS) conditions, TOA measurements collected at spatially distributed receivers form a set of nonlinear equations whose solution defines the target position. The inherent nonlinearity and non-convexity of this problem make classical gradient-based solvers susceptible to local minima, motivating the use of hybrid metaheuristic optimization approaches. This paper investigates the application of hybrid optimization algorithms combining genetic algorithm (GA) with local search refinement for solving the passive TOA localization problem. Performance of considered algorithms is evaluated using root mean square error (RMSE) and compared against the Cramer-Rao lower bound (CRLB) as a theoretical benchmark. Results demonstrate that the hybrid approach consistently outperforms classical nonlinear least squares methods and conventional optimization algorithms in both accuracy and robustness, with hybrid optimization algorithm showing superior convergence in high-noise scenarios.*

Keywords: *passive localization, TOA, Genetic Algorithm, Hybrid optimization*

1. INTRODUCTION

The problem of determining the unknown position of a target from measurements collected at a set of spatially distributed receivers with known locations is a fundamental task in numerous military and civilian applications, including radar and sonar surveillance, electronic warfare, target tracking, navigation and wireless sensor networks [1], [2] .In passive localization the target does not cooperate with the localization system, and its position has to be inferred solely from the signals intercepted by the receivers, which makes the problem particularly relevant for modern defense and surveillance systems. The key requirement common to all of these applications is to estimate the target location as accurately as possible from a set of noisy measurements.

Depending on the available hardware, localization systems exploit different measurement types, such as the time of arrival (TOA), the time difference of arrival (TDOA), the received signal strength (RSS) or the angle of arrival (AOA) [3], [4]. This paper focuses on TOA-based localization, in which the position of the target is determined from the propagation times of the emitted signal to the receivers. Assuming synchronized receivers, each TOA measurement is converted into a range between the target and the corresponding receiver, defining a circle on which the target lies. TOA measurements are attractive because of their high ranging accuracy and the relatively simple required hardware [5].

Because the measured ranges are corrupted by noise, the localization task is naturally formulated as an optimization problem based on the least squares (LS) method, in which the sum of squared differences between the measured and the modelled ranges is minimized. LS estimators are commonly divided into linear least squares (LLS), which provide a closed-form solution after a suitable linearization, and nonlinear least squares (NLS), which operate directly on the original nonlinear model [6]. The weighted least squares (WLS) variant assigns to each residual a weight inversely proportional to its variance and yields an efficient closed-form estimate. However, the objective function of the NLS estimator is highly nonlinear and multimodal, so that conventional gradient-based optimization algorithms strongly depend on the initial solution and may converge to a local rather than to the global minimum.

To overcome this difficulty, researchers have proposed different hybrid variations of evolutionary algorithms [7]. Therefore, this paper proposes a hybrid metaheuristic optimization approach that combines a Genetic Algorithm (GA) with the Nelder-Mead (NM) local search method. The GA is a population-based stochastic optimizer inspired by natural selection [8] that explores the search space globally and is able to escape local minima, whereas the NM simplex method [9] efficiently refines a candidate solution without requiring derivative

Corresponding author: M. Rosić;
E-mail: mrosic@mas.bg.ac.rs

DOI: 10.5937/OTEH26087R

information. By hybridizing the two algorithms, the global exploration capability of the GA and the local exploitation efficiency of the NM method are combined into a single robust estimator, denoted GA-NM. The Cramer-Rao lower bound (CRLB) [10] is derived for the considered TOA model and used as a theoretical benchmark, while the root mean square error (RMSE) is adopted as the accuracy measure. The performance of the proposed GA-NM method is compared with the closed-form WLS estimator and with the standalone GA.

The remainder of the paper is organized as follows. Section 2 formulates the passive TOA localization problem. Section 3 presents the NLS and WLS formulations. Section 4 describes the proposed hybrid GA-NM method. Section 5 derives the CRLB for TOA measurements. Section 6 reports and discusses the simulation results, and Section 7 concludes the paper.

2. PROBLEM FORMULATION

Consider the two-dimensional (2D) passive localization of a single target under line-of-sight (LOS) conditions. The considered problem is consisting of a set of receivers $\mathbf{x}_i^r = \begin{bmatrix} x_i^r & y_i^r \end{bmatrix}^T \in \mathbb{R}^2, \forall i \in \{1,2,...,N\}$ with known coordinates, one transmitter $\mathbf{x}_t = \begin{bmatrix} x_t & y_t \end{bmatrix}^T$, and one passive target at unknown position $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T \in \mathbb{R}^2$. Without the loss of generality, the transmitter is placed in the origin, i.e., $\mathbf{x}_t = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, as illustrated in Fig. 1. To determine the target position from passive TOA measurements, at least three receivers (N ≥ 3) are required.

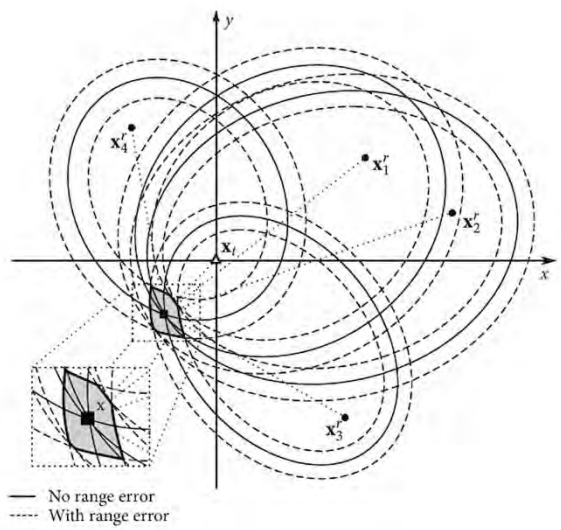


Figure 1. Geometrical model based on TOA.

In a passive target localization system, a signal is first transmitted and then reflected by the target upon arrival. The Time of Arrival (TOA) method considered in this work determines range measurements by exploiting the absolute travel time of the signal between the transmitter and the receivers. Assuming synchronized receivers, the range between the target and the i-th receiver is obtained by multiplying the measured propagation time by the speed of light c, as given in Eq. (1).

$$r_i = \frac{1}{c}\left(\|\mathbf{x}\|_2 + \|\mathbf{x} - \mathbf{x}_i^r\|_2\right) + \tilde{n}_i, \quad i = 1,2,...,N \tag{1}$$

where $\|\square\|_2$ denotes the Euclidian distance. In the absence of noise, this range equals the Euclidean distance between the target and the receiver, Eq. (2).

$$d_i = \sqrt{x^2 + y^2} + \sqrt{\left(x - x_i^r\right)^2 + \left(y - y_i^r\right)^2} \tag{2}$$

In practice the measurements are corrupted by noise, which is modelled as a set of independent zero-mean Gaussian random variables $\tilde{n}_i$ with variance σ^2. The measured range is therefore expressed by Eq. (3).

$$r_i = d_i + \tilde{n}_i \tag{3}$$

Each noiseless TOA measurement constrains the target to lie on a ellipse with foci at the corresponding receiver and transmitter. As shown in Fig. 1, in the absence of noise the target position is uniquely determined by the intersection of at least three such ellipses. When noise is present, the ellipses no longer intersect at a single point, and an appropriate optimization technique has to be employed in order to minimize the localization error.

3. LEAST SQUARE METHODS

Following the TOA measurement model, the localization problem is formulated as an NLS problem. The objective function $J_{NLS}(\mathbf{x})$ is defined as the sum of squared residuals between the measured and the modelled ranges, Eq. (4).

$$J_{NLS}(\mathbf{x}) = \sum_{i=1}^{N}\left(d_i - \|\mathbf{x} - \mathbf{x}_i^r\|\right)^2 \tag{4}$$

The optimal estimate of the target position is then obtained as the minimizer of this objective function over the feasible region, Eq. (5). Because the modelled range is a nonlinear function of the target coordinates, the objective function is non-convex and may exhibit multiple local minima.

$$\hat{x} = \arg\min_{\mathbf{x} \in \mathbb{R}^2} J_{NLS}(\mathbf{x}) \tag{5}$$

In order to obtain a closed-form estimate, the nonlinear model is linearized. Squaring the range equation yields the following

$$d_i^2 = \|\mathbf{x}\|^2 - 2\mathbf{x}_i^{rT}\mathbf{x} + \|\mathbf{x}_i^r\|^2 \tag{6}$$

Subtracting the equation of the reference receiver ($i = 1$) eliminates this common term and results in the linear relation given by Eq. (7).

$$\begin{aligned} 2x_i^r x + 2y_i^r y - 2r_i R^t = \left(x_i^r\right)^2 + \left(y_i^r\right)^2 - r_i^2 \\ -2n_i R^t + 2r_i n_i \end{aligned} \tag{7}$$

where $R^t = \|\mathbf{x}\|_2$. Collecting the equations for i = 2, ..., N leads to the matrix form of Eq. (8), where the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are defined in Eq. (9) and Eq. (11), respectively.

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b} + \mathbf{m}, \tag{8}$$

$$\mathbf{A} = 2\begin{bmatrix} x_1^r & y_1^r & -r_1 \\ \vdots & \vdots & \vdots \\ x_N^r & y_N^r & -r_N \end{bmatrix}, \tag{9}$$

$$\boldsymbol{\theta} = \begin{bmatrix} x & y & R^t \end{bmatrix}^T, \tag{10}$$

$$\mathbf{b} = \begin{bmatrix} (x_1^r)^2 + (y_1^r)^2 - r_1^2 \\ \vdots \\ (x_N^r)^2 + (y_N^r)^2 - r_N^2 \end{bmatrix}, \tag{11}$$

$$\mathbf{m} = 2\begin{bmatrix} (r_1 - R^t)n_1 & \cdots & (r_N - R^t)n_N \end{bmatrix}^T. \tag{12}$$

Based on the linear model, the WLS objective function is defined in Eq. (13), where $\mathbf{W}$ is a weighting matrix chosen as the inverse of the covariance matrix of the linearized measurement errors.

$$J_{WLS}(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{W} (\mathbf{A}\mathbf{x} - \mathbf{b}) \tag{13}$$

The corresponding closed-form WLS estimate is given by Eq. (14). The main advantages of the WLS estimator are its low computational cost and the absence of an initialization requirement. Nevertheless, the linearization introduces a bias, and the estimator does not attain sufficient accuracy in highly nonlinear, low-SNR scenarios, which motivates the use of the nonlinear formulation solved by metaheuristic optimization.

$$\hat{\mathbf{x}}_{WLS} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b} \tag{14}$$

4. HYBRID GA-NM METHOD

The proposed hybrid GA-NM method is described in this section in the context of passive TOA localization. The main idea behind hybridizing two optimization algorithms is to combine the advantages of each while mitigating their drawbacks. The GA is a powerful global optimizer that explores the search space by evolving a population of candidate solutions, whereas the NM method is an efficient local search technique that rapidly refines a given solution. In the proposed approach, the GA first identifies the region of the global optimum, after which the NM method fine-tunes the corresponding solution.

4.1. Genetic algorithm

The GA is a metaheuristic optimization method based on the mechanisms of natural selection and genetics [8]. To apply the GA to the NLS problem, the minimization in Eq. (4) is augmented with bound constraints, Eq. (15), where $\mathbf{x}^l$ and $\mathbf{x}^h$ denote the lower and upper bounds of the search region. The bound constraints prevent the objective function from being evaluated at infeasible points.

$$\min_{\mathbf{x}^l \le \mathbf{x} \le \mathbf{x}^h} J_{NLS}(\mathbf{x}) \tag{15}$$

The algorithm starts from a population of N_P individuals randomly generated within the feasible region, each representing a candidate target position, and evolves it through three genetic operators: selection, crossover and mutation. In the selection step, individuals are chosen for reproduction according to their fitness. Using roulette-wheel selection, the probability of selecting the i-th individual is proportional to its fitness, Eq. (16), so that individuals associated with lower objective-function values are more likely to be selected.

$$P_i = \frac{F_i}{\sum_{j=1}^{N_p} F_j}, \quad i = 1, \ldots, N_p \tag{16}$$

The crossover operator combines the information of two selected parents to produce offspring that inherit favorable traits from both, thereby driving the search towards promising regions. The mutation operator introduces small random perturbations into a limited number of individuals, maintaining population diversity and preventing premature convergence to local minima. The evolutionary process is repeated until a termination criterion is satisfied, for example a maximum number of generations or a sufficiently small relative change of the average population fitness between two consecutive generations, Eq. (17).

$$\left| \frac{f_{avg}^{(k)} - f_{avg}^{(k-1)}}{f_{avg}^{(k)}} \right| \le \varepsilon \tag{17}$$

4.2. Nelder-Mead method

The NM simplex method is a derivative-free local search technique for minimizing a function in an n-dimensional space [9]. It operates on a simplex of $n+1$ vertices, which for the considered 2D problem is a triangle. At each iteration the vertices are ordered according to their objective-function values, and the simplex is modified by means of four geometric operations: reflection, expansion, contraction and shrinkage. The reflection of the worst vertex through the centroid of the remaining vertices is computed by Eq. (18).

$$\mathbf{x}_r = \bar{\mathbf{x}} + \rho\left(\bar{\mathbf{x}} - \mathbf{x}_{n+1}\right) \tag{18}$$

If the reflected point improves the best value, the simplex is expanded according to Eq. (19); if the reflection does not yield an improvement, the simplex is contracted, Eq. (20); and if contraction fails, the simplex is shrunk towards the best vertex, Eq. (21). Here ρ, χ, γ and δ denote the reflection, expansion, contraction and shrinkage coefficients, respectively.

$$\mathbf{x}_e = \bar{\mathbf{x}} + \chi\left(\mathbf{x}_r - \bar{\mathbf{x}}\right) \tag{19}$$

$$\mathbf{x}_c = \bar{\mathbf{x}} + \gamma\left(\mathbf{x}_{n-1} - \bar{\mathbf{x}}\right) \tag{20}$$

$$\mathbf{x}_i = \mathbf{x}_1 + \delta(\mathbf{x}_i - \mathbf{x}_1), \quad i = 2,\ldots,n+1 \tag{21}$$

The process is repeated until the size of the simplex falls below a prescribed tolerance. The geometric illustration on 2D problem of different operations of the NM algorithm are shown in Figure 2.

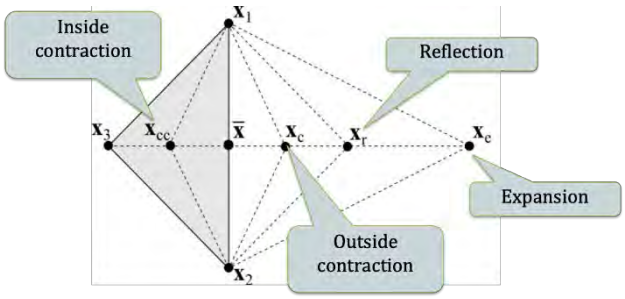


Figure 2. Geometrical illustration of NM method

4.3. Proposed hybrid GA-NM procedure

The proposed hybrid GA-NM method operates in two phases. In the global search phase, the GA evolves an initial population within the bound-constrained search region and, after a predefined number of generations, returns the best individual as an estimate of the region containing the global optimum. In the local refinement phase, this best individual is used to construct the initial simplex of the NM method, which then refines the solution through the geometric operations described above. If the refined solution does not satisfy the prescribed accuracy, a new GA cycle is initiated; otherwise the algorithm terminates and returns the final target-position estimate. By alternating between global exploration and local exploitation, the hybrid GA-NM method avoids the local minima that hamper purely local solvers and overcomes the limited final accuracy of the standalone GA, providing precise and robust localization even in noisy conditions. The pseudocode of the proposed hybrid GA-NM algorithm is shown in Figure 3.

5. CRAMER-RAO LOWER BOUND

The CRLB provides a theoretical lower bound on the variance of any unbiased estimator and is widely used to assess the efficiency of localization algorithms [10]. It is obtained from the inverse of the Fisher information matrix (FIM), and the position RMSE bound is given by the square root of the trace of the inverse FIM, Eq. (22).

$$\mathrm{CRLB}(\mathbf{x}) = \mathrm{trace}\left\{\mathbf{I}^{-1}(\mathbf{x})\right\} \tag{22}$$

For the considered passive TOA model with independent Gaussian range errors of variance σ^2, the FIM reduces to

$$\mathbf{I}(\mathbf{x}) = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}, \tag{23}$$

where the appropriate components are

$$\begin{aligned} I_{xx} &= \sum_{i=1}^{N} \frac{1}{\sigma_{ni}^2}\left(\frac{x - x_i^r}{\|\mathbf{x}-\mathbf{x}_i^r\|_2} + \frac{x}{\|\mathbf{x}\|_2}\right)^2, \\ I_{xy} = I_{yx} &= \sum_{i=1}^{N} \frac{1}{\sigma_{ni}^2}\left(\frac{x - x_i^r}{\|\mathbf{x}-\mathbf{x}_i^r\|_2} + \frac{x}{\|\mathbf{x}\|_2}\right) \\ &\quad \times\left(\frac{y - y_i^r}{\|\mathbf{x}-\mathbf{x}_i^r\|_2} + \frac{y}{\|\mathbf{x}\|_2}\right), \\ I_{yy} &= \sum_{i=1}^{N} \frac{1}{\sigma_{ni}^2}\left(\frac{y - y_i^r}{\|\mathbf{x}-\mathbf{x}_i^r\|_2} + \frac{y}{\|\mathbf{x}\|_2}\right)^2. \end{aligned} \tag{24}$$

The CRLB therefore depends both on the measurement noise level and on the geometric configuration of the receivers relative to the target.

Algorithm 1 Hybrid GA–NM Method for Target Localization

Require: Search bounds $[\mathbf{x}^l, \mathbf{x}^h]$, population size N_p, max generations $G_{\max}$, NM coefficients $\rho, \chi, \gamma, \delta$, accuracy threshold ε, max restarts $K_{\max}$
Ensure: Target position estimate $\hat{\mathbf{x}}$
Initialize restart counter $k \leftarrow 0$
Initialize best solution $\hat{\mathbf{x}} \leftarrow \emptyset$
repeat
  $k \leftarrow k + 1$ ▷ *Phase 1: Global Search via GA*
  Initialize population $\mathcal{P} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_p}\}$ uniformly within $[\mathbf{x}^l, \mathbf{x}^h]$
  for $g = 1$ **to** $G_{\max}$ **do**
    Evaluate objective $J_{NLS}(\mathbf{x}_i)$ for all $\mathbf{x}_i \in \mathcal{P}$
    Select parents via roulette-wheel selection ▷ Eq. (14)
    Apply crossover and mutation operators to generate offspring
    Replace population $\mathcal{P}$ with new generation
  end for
  $\mathbf{x}^* \leftarrow \arg\min_{\mathbf{x}_i \in \mathcal{P}} J_{NLS}(\mathbf{x}_i)$
  ▷ *Phase 2: Local Refinement via Nelder–Mead*
  Construct initial simplex $\mathcal{S}$ centered at $\mathbf{x}^*$
  repeat
    Order simplex vertices:
    $J_{NLS}(\mathbf{x}_1) \leq \cdots \leq J_{NLS}(\mathbf{x}_{n+1})$
    Compute centroid $\bar{\mathbf{x}}$ of all vertices except $\mathbf{x}_{n+1}$
    Attempt **reflection**: $\mathbf{x}_r$
    if $J_{NLS}(\mathbf{x}_r) < J_{NLS}(\mathbf{x}_1)$ **then**
      Attempt **expansion**: $\mathbf{x}_e$
      Accept $\mathbf{x}_e$ or $\mathbf{x}_r$ (whichever is better)
    else if $J_{NLS}(\mathbf{x}_r) < J_{NLS}(\mathbf{x}_n)$ **then**
      Accept reflection: replace $\mathbf{x}_{n+1} \leftarrow \mathbf{x}_r$
    else
      Attempt **contraction**: $\mathbf{x}_c$
      if $J_{NLS}(\mathbf{x}_c) < J_{NLS}(\mathbf{x}_{n+1})$ **then**
        Accept contraction: replace $\mathbf{x}_{n+1} \leftarrow \mathbf{x}_c$
      else
        Apply **shrink**:
        $\mathbf{x}_i \leftarrow \mathbf{x}_1 + \delta(\mathbf{x}_i - \mathbf{x}_1)$, $i = 2, \ldots, n+1$
      end if
    end if
  until NM convergence criterion satisfied
  $\hat{\mathbf{x}} \leftarrow \mathbf{x}_1$
until $J_{NLS}(\hat{\mathbf{x}}) \leq \varepsilon$ **or** $k \geq K_{\max}$
return $\hat{\mathbf{x}}$

Figure 3. Pseudocode of the GA-NM method

6. SIMULATION RESULTS

This section presents the results of numerical simulations carried out to evaluate the localization performance of the proposed hybrid GA-NM method against the closed-form WLS estimator, the standalone GA, and the derived CRLB. The considered scenario comprises five receivers placed at the known coordinates [200, 300], [1500, 100], [120, 1500], [1510, 1500] and [700, 900] m, while the true target position is set to [300, 500] m. The GA uses a population of 40 individuals, a crossover probability of 0.9 and a mutation probability of 0.12, while the relative measurement-noise level is controlled through the signal-to-noise ratio (SNR), with the noise standard deviation given by $\sigma = \sigma_0 \cdot 10^{(-SNR/20)}$. The localization accuracy is quantified by the RMSE, Eq. (25), evaluated over M = 1000 independent Monte Carlo runs, where $\hat{x}(n)$ denotes the estimate obtained in the n-th run.

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{n=1}^{M}\|\hat{\mathbf{x}}(n) - \mathbf{x}\|^2} \tag{25}$$

Fig. 4 illustrates the convergence behaviour of the GA for a representative noisy realization, showing the best objective-function value as a function of the generation number. The objective value decreases rapidly during the first generations, as the population concentrates around the global optimum, and subsequently flattens. This confirms that the GA quickly locates the region of the global minimum but exhibits a slow final convergence, which motivates the local refinement performed by the NM method in the proposed hybrid scheme.

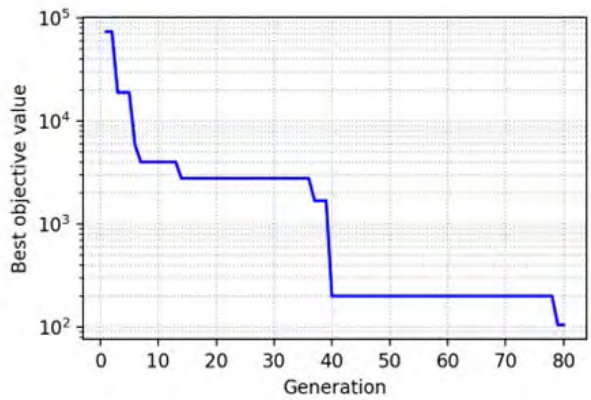


Figure 4. Convergence of the genetic algorithm for a representative noisy realization.

Fig. 5 shows the RMSE of the considered algorithms, expressed as 10 log(RMSE), as a function of the SNR. The proposed hybrid GA-NM method attains the CRLB over the entire range of SNR values, demonstrating its accuracy and robustness in different noise conditions. The closed-form WLS estimator follows the CRLB closely but remains slightly above it, particularly in the low-SNR region, where the linearization bias is more pronounced. In contrast, the standalone GA reaches the global region but is unable to refine the solution to the required precision, exhibiting an error floor at moderate and high SNR. This clearly demonstrates that the local refinement provided by the NM method is essential for achieving CRLB-level accuracy.

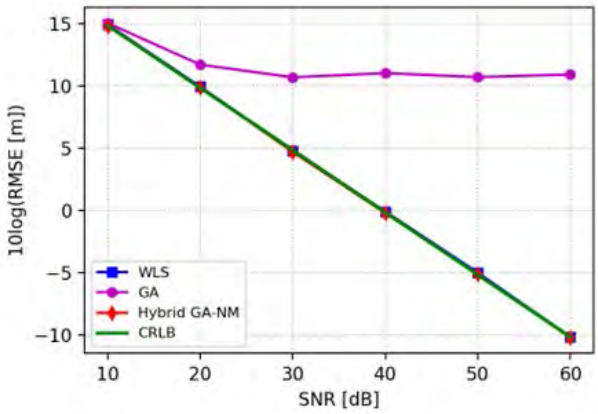


Figure 5. Comparison of RMSE versus SNR levels for the considered algorithms.

To further assess the localization performance, the cumulative distribution functions (CDFs) of the localization error are compared at two SNR levels, 10 dB and 50 dB, in Figs. 6 and 7, respectively. At SNR = 10 dB (Fig. 4), the hybrid GA-NM method achieves the highest CDF values across the whole error range, indicating that it produces small localization errors more frequently than the other methods, while the WLS estimator performs only marginally worse and the standalone GA exhibits a slightly heavier tail.

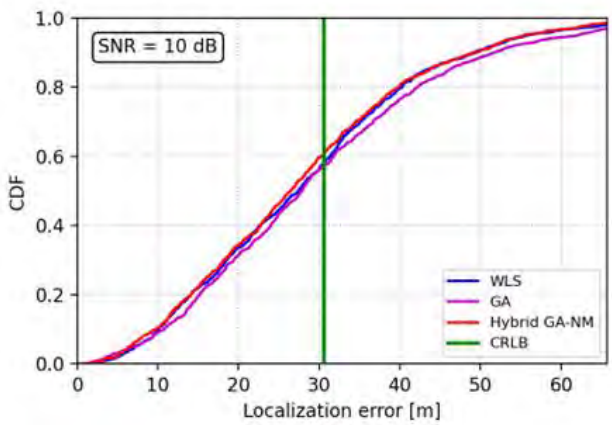


Figure 6. CDFs of the localization error of different algorithms for SNR = 10 dB.

At SNR = 50 dB (Fig. 7), both the hybrid GA-NM and the WLS estimators yield sub-meter localization errors that cluster around the CRLB, whereas the error of the standalone GA lies well outside the displayed sub-meter range, in agreement with the error floor observed in Fig. 5. These results confirm that the proposed hybrid GA-NM method offers superior accuracy and robustness compared with the conventional WLS estimator and the standalone GA.

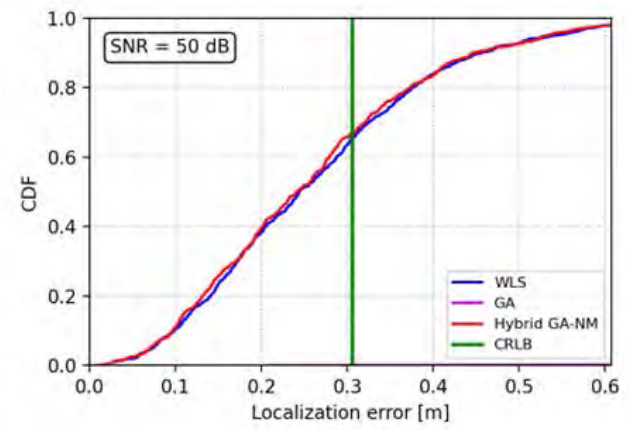


Figure 7. CDFs of the localization error of different algorithms for SNR = 50 dB.

7. CONCLUSION

This paper has addressed the passive TOA-based localization of a target from range measurements collected at several receivers with known positions. The localization problem has been formulated as a nonlinear least squares optimization problem, and a hybrid metaheuristic method combining a Genetic Algorithm with the Nelder-Mead local search has been proposed to solve it. The GA performs the global exploration of the multimodal objective function, while the NM method efficiently refines the resulting solution, so that the hybrid GA-NM estimator combines the advantages of both techniques.

The CRLB has been derived for the considered TOA model and used, together with the RMSE and the CDF of the localization error, to evaluate the performance of the proposed method. The simulation results show that the hybrid GA-NM method attains the CRLB over a wide range of SNR values and consistently outperforms both the standalone GA, which suffers from a limited final accuracy, and the closed-form WLS estimator, whose linearization bias degrades the accuracy at low SNR. The proposed method therefore provides accurate and robust passive localization in noisy environments.

Future work will focus on extending the proposed approach to three-dimensional localization, to scenarios involving multiple targets and non-line-of-sight propagation, and on incorporating adaptive strategies for tuning the parameters of the metaheuristic algorithm.

Acknowledgment

The research of M. Rosić was supported by the Serbian Ministry of Education, Science and Technological Development under Grant TR35029. The research of M. Sedak was supported by the Serbian Ministry of Education, Science and Technological Development under Grant No. TR35006.

References

[1] J. Figueiras and S. Frattasi, *Mobile Positioning and Tracking*. United Kingdom: John Wiley & Sons, 2010.

[2] R. Zekavat and R. M. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*. John Wiley & Sons, 2011.

[3] I. Guvenc and C.-C. Chong, "A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques," *IEEE Commun. Surv. Tutor.*, vol. 11, no. 3, pp. 107–124, 2009.

[4] V. R. Kulkarni, V. Desai, and R. V. Kulkarni, "A Comparative Investigation of Deterministic and Metaheuristic Algorithms for Node Localization in Wireless Sensor Networks," *Wirel. Netw.*, vol. 25, no. 5, pp. 2789–2803, 2019.

[5] Y. T. Chan and K. C. Ho, "A Simple and Efficient Estimator for Hyperbolic Location," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1905–1915, 1994.

[6] B. Chalise, Y. Zhang, M. G. Amin, and B. Himed, "Target Localization in a Multi-static Passive Radar System through Convex Optimization," *Signal Process.*, vol. 102, pp. 207–215, 2014.

[7] M. B. Rosić, M. Sedak, M. Simić, and P. Pejović, "An Improved Chaos Driven Hybrid Differential Evolution and Butterfly Optimization Algorithm for Passive Target Localization Using TDOA Measurements," *Appl. Sci.*, vol. 13, no. 2, p. 684, 2023.

[8] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[9] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[10] M. Laaraiedh, S. Avrillon, and B. Uguen, "Cramer-Rao Lower Bounds for Nonhybrid and Hybrid Localisation Techniques in Wireless Networks," *Eur. Trans. Telecommun.*, vol. 23, pp. 268–280, 2012.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EFFICIENCY ANALYSIS OF A STEGANOGRAPHY METHOD BASED ON PIXEL VALUE DIFFERENCES

SANJA KLAJIĆ
Center for Applied Mathematics and Electronics, Belgrade, sanja.klajic@vs.rs

BRANKO VUJATOVIĆ
Center for Applied Mathematics and Electronics, Belgrade, branko.vujatovic@vs.rs

DARKO GRUBAČ
Center for Applied Mathematics and Electronics, Belgrade, darko.grubac@vs.rs

NENAD STOJANOVIĆ
University of Defence, Military Academy, and
Center for Applied Mathematics and Electronics, Belgrade, nenad.m.stojanovic@vs.rs

Abstract: *This paper describes in detail a steganography method that embeds information into the difference between neighboring pixel values. The efficiency analysis of the method was performed through the aspects of imperceptibility, security, robustness, capacity, and complexity. Imperceptibility was analyzed using objective image quality assessment methods. For security analysis, statistical imperceptibility based on histogram were employed. Robustness was analyzed through the accuracy of extracting the embedded message at a given embedding capacity and under the influence of additive noise of different intensities. Capacity was measured depending on the characteristics of the cover image, while the complexity of the hiding algorithm was determined by measuring the information embedding execution time. The obtained results were compared with the characteristics of a steganography method based on least significant bit substitution. It was shown that neither of the two methods demonstrated superior efficiency across all considered criteria. Analysis of the obtained results for all cover images showed that imperceptibility and security favored the method based on pixel value differences, while robustness, capacity, and complexity favored the method based on least significant bit substitution. The method that should be used in practical applications depends on the requirements of the transmission system. It is necessary to find a compromise between the advantages and disadvantages of the considered methods, especially since the differences in performance for individual characteristics may not necessarily be significant. This requires an appropriate selection of the cover image, which is an important factor in steganography systems.*

Keywords: *pixel value differences, cover image, imperceptibility, capacity*

1. INTRODUCTION

In the modern digital environment, the rapid expansion of the Internet and the growing exchange of sensitive data have significantly increased the demand for secure communication. In this context, steganography and cryptography play a crucial role in information protection. Cryptography techniques rely on encrypting data and transmitting it in an unreadable form, which may draw the attention of unauthorized users [1]. In contrast, steganography aims to conceal the very existence of a secret message by embedding data into appropriate cover media, thereby reducing the likelihood of attracting attention and revealing the communication [2]. The combined application of cryptography and steganography can further enhance the level of data security.

The term steganography originates from the Greek words stegos, meaning to "to conceal" or "covered", and graphia, meaning "writing". Thus, the term literally translates as "concealed writing".

The origins of steganography date back to ancient times, when secret messages were concealed using various methods, such as tattooing or invisible ink. With the development of digital technologies, steganography has undergone a significant transformation and is now primarily based on embedding information within various types of digital cover media [3].

The cover media used to conceal secret messages or files may include various types of digital content, such as images, audio recordings, video sequences, and text documents. Among these, digital images are particularly well suited for steganography applications due to their widespread use and their ability to accommodate imperceptible modifications [4]. In this study, a digital image is selected as the cover medium for embedding hidden information.

The key performance aspects of steganography include imperceptibility, security, robustness, embedding capacity, and computational complexity. An important

Corresponding author: S. Klajić;
E-mail: sanja.klajic@vs.rs
DOI: 10.5937/OTEH26088K

question is which of these aspects is the most significant when evaluating the effectiveness of different steganography techniques [5]. Since it is not possible to simultaneously optimize all of these performance criteria, an appropriate trade-off must be established among them. Consequently, the suitability of a particular steganography method depends on the requirements and constraints of the intended application.

The objective of this study is to provide a comprehensive analysis of the Pixel Value Differencing (PVD) method and to compare its performance characteristics with those of the Two Least Significant Bits (2-LSB) method in order to identify their respective advantages and limitations.

The remainder of this paper is organized as follows. Section 2 describes the methods and techniques employed in the conducted study. Section 3 presents and discusses the obtained results, while Section 4 summarizes the main conclusions.

2. METHODOLOGY

2.1. Characteristics of the Cover Images and the Secret Message

For the purposes of this study, a dataset consisting of 50 images obtained from a publicly available image database was used [6]. All selected images have a resolution of 1920×1080 pixels and were employed as cover images for embedding the secret message. Prior to the embedding process, all images were converted to grayscale, in which each pixel is represented by a single 8-bit intensity value within the range [0-255].

The methods were compared from the standpoint of maximum embedding capacity. A fixed binary sequence of 4,147,200 bits was used as the secret message, corresponding to the maximum embedding capacity of the 2-LSB method for a 1920×1080 grayscale image. The same payload was applied to the PVD method, whose embedding capacity depends on the characteristics of each individual cover image.

2.2. Description of the PVD Method

In this study, the PVD method proposed in [7] was employed. In this method, data are embedded into the differences between the intensity values of neighboring pixels, enabling secret information to be concealed within the cover image. The magnitude of the pixel-value difference determines the amount of data that can be embedded. Small differences between adjacent pixel values correspond to smooth regions of the image and therefore allow only a limited amount of information to be embedded. In contrast, larger differences, typically found in high-contrast regions or along edges, provide a higher embedding capacity while maintaining the imperceptibility of the embedded information.

In this study, an approach based on forming pairs of adjacent pixels along the image rows is adopted. First, the image is partitioned into pairs of neighboring, non-overlapping pixels. For each pixel pair, denoted by p_i and p_{i+1}, the difference is calculated as:

$$d = p_{i+1} - p_i, \tag{1}$$

as well as its absolute value $|d|$, which can take values in the range 0 to 255. The absolute values of the differences are classified into predefined ranges R_k, where $k = 1, 2, ..., n$. The lower and upper bounds of the k-th range are denoted by l_k i u_k, respectively. The minimum value of the first range is $l_1 = 0$, while the maximum value of the last range is $u_n = 255$. The width of the k-th range, denoted by W_k is determined by the following expression:

$$W_k = u_k - l_k + 1. \tag{2}$$

Here, the width of each range is chosen as a power of two. Accordingly, the following ranges are defined: [0-7], [8-15], [16-31], [32-63], [64-127], and [128-255]. Consequently, ranges corresponding to small pixel-value differences have narrow widths and represent smooth image regions. In contrast, ranges associated with large differences are assigned wider intervals and correspond to edge regions. The purpose of this range selection is to achieve greater visual imperceptibility. The number of bits embedded into the difference of each pixel pair depends on the range to which the difference belongs. It is denoted by t and determined according to the following expression:

$$t = \log_2(W_k). \tag{3}$$

Then, a substring consisting of t bits is extracted from the secret message. This binary substring is converted to the corresponding decimal value, denoted by b. New difference d' is calculated as

$$d' = \begin{cases} l_k + b, & d \geq 0 \\ -(l_k + b), & d < 0 \end{cases} \tag{4}$$

The goal is to compute new pixel values so that their difference is exactly equal d'. After determining the new difference d', the value m is calculated, which represents the difference between the new and original pixel difference:

$$m = d' - d. \tag{5}$$

Based on the value of m, the modification of pixel values is performed and depends on the parity of the original difference d. If the value $m/2$ is not an integer, rounding is applied so that one part of the change is rounded down to a smaller integer and the other to a larger integer. This ensures that the new difference between the modified pixels is equal to the value d', while the total change is distributed almost evenly between the pixels. This reduces image distortion, achieving a higher level of imperceptibility. The new pixel values are obtained by applying the following relations:

$$\left(p_i', p_{i+1}'\right) = \begin{cases} p_i - \left\lceil \frac{m}{2} \right\rceil,\ p_{i+1} + \left\lfloor \frac{m}{2} \right\rfloor, & \text{if } d \text{ is odd} \\ p_i - \left\lfloor \frac{m}{2} \right\rfloor,\ p_{i+1} + \left\lceil \frac{m}{2} \right\rceil, & \text{if } d \text{ is even} \end{cases} \tag{6}$$

During the computation of the new pixel values, it is possible for one or both values to fall outside the valid

intensity range. In the original approach, problematic blocks are excluded from the embedding process [7]. In contrast, some modifications of the PVD method employ various correction procedures that involve adjusting pixel values to bring them back into the allowable range [1].

In this work, pixel-value shifting correction is applied, in which both pixels are shifted by the same amount to bring their values back into the permissible range. In this way, the difference between their values remains unchanged. If one of the values is less than 0, both pixel values are increased by the corresponding amount. If one of the values exceeds 255, both pixel values are decreased by the amount of the overflow. This procedure avoids discarding pixel pairs while preserving the difference that carries the embedded information. The correction can be formally expressed as:

$$\Delta = \begin{cases} -\min(p_i', p_{i+1}'), & \text{if } \min(p_i', p_{i+1}') < 0, \\ 255 - \max(p_i', p_{i+1}'), & \text{if } \max(p_i', p_{i+1}') > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

The corrected pixel values are then given by:

$$p_i'' = p_i' + \Delta, \quad (8)$$

$$p_{i+1}'' = p_{i+1}' + \Delta. \quad (9)$$

The extraction process follows the same image processing principle as the secret message embedding process, processing the image sequentially in pairs of pixels. The absolute value of the difference d'' between adjacent pixels of the stego image is calculated, determining the range to which it belongs and the number of bits embedded in the observed pair of pixels. The value b, that is, the part of the message that was inserted into the corresponding pair of pixels, is extracted based on the following expression:

$$b = d'' - l_k. \quad (10)$$

When extracting a secret message from a stego image, it is not necessary to use the original image for comparison.

The following example illustrates the secret message embedding process. Let $(p_i, p_{i+1}) = (100, 70)$ be a pair of cover image pixels. The difference between the pixels is

$$d = p_{i+1} - p_i = 70 - 100 = -30.$$

The absolute value of the difference $|d| = 30$ belongs to the interval [16-31]. The width of the interval is:

$$W_k = u_k - l_k + 1 = 31 - 16 + 1 = 16,$$

so the number of bits that can be embedded is equal to:

$$t = log_2 16 = 4.$$

Assume that 1100 are the next four bits of the message to be embedded. Their decimal value is $b = 12$. Since $d < 0$, then:

$$d' = -(l_k + b) = -(16 + 12) = -28.$$

Therefore, the new difference is:

$$m = d' - d = -28 - (-30) = 2.$$

Since |d|=30 is even, the resulting pair of stego pixels is obtained according to Equation (6): $(p_i', p_{i+1}') = (99, 71)$.

During the extraction process, the resulting pair of stego pixels is used $(p_i', p_{i+1}') = (99, 71)$. The absolute value of the difference between the stego pixels is $|d''| = 28 \in$ [16-31]. The number of extracted bits is determined according to the previously presented equations. Therefore:

$$W_k = u_k - l_k + 1 = 31 - 16 + 1 = 16,$$

$$t = \log_2 16 = 4.$$

The decimal value of the embedded bits is:

$$b = |d''| - l_k = 28 - 16 = 12,$$

which, converted into a binary record of 4 bits length, gives 1100.

2.3. 2-LSB Method

The LSB method is one of the fundamental and most commonly used techniques in spatial domain steganography. This method involves replacing the least significant bits of the cover image pixel with the corresponding bits of the secret message [8]. It is a basic method and most often is part of hybrid methods, as well as in implementations of message hiding in transform domains, such as the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT). In order to compare the characteristics with the PVD method, a 2-LSB steganography method was implemented. In this variant of the LSB method, 2 bits of the secret message are embedded in each pixel. In previous comparisons, the case where the LSB method changes exactly one bit (the least significant one) has been the primary focus [9].

2.4. Measurement Methods

In order to evaluate the effectiveness of information hiding from the aspects of imperceptibility, robustness, security, capacity, and complexity, appropriate objective metrics were used.

The imperceptibility of the embedded secret message was evaluated using the following objective image quality assessment metrics: Peak Signal-to-Noise Ratio (PSNR) [10], Structural Similarity Index Metric (SSIM) [11], and Additive Impairments and Detail Losses Metric (ADM) [12]. These metrics were used to determine the degree of similarity between the cover image and the stego image.

The robustness was analyzed against Additive White Gaussian Noise (AWGN) in the context of a potential jamming attack. Robustness assessment was performed using the Bit Error Rate (BER). The metric represents the ratio between the number of incorrectly extracted bits and the total number of embedded bits, thereby evaluating the success of the hidden message reconstruction.

The security of the analyzed methods is confirmed by statistical imperceptibility. For analysis purposes, the histograms of the cover and stego images were used. The degree of histogram similarity was determined using the histogram correlation metric and the Root Mean Square Error (RMSE). Based on the obtained values, the extent of changes in the pixel intensity distribution introduced during the embedding process was assessed. In addition to

histogram-based analysis, security should also be considered from the aspect of steganalysis. As the considered methods are fundamentally different, different steganalysis techniques will show greater efficiency on the method for which they were originally developed [13].

The computational complexity of the algorithm was estimated based on the time required to embed the secret message into the cover image. The execution time was measured using corresponding MATLAB functions on a computer equipped with an Intel Core i7-1255U processor, 16 GB of RAM, and the Microsoft Windows 10 Pro operating system.

The capacity of a method represents the maximum amount of data that can be embedded in a cover image. It is determined by calculating the number of bits that can be embedded in the pixels of the cover image.

3. RESULTS ANALYSIS

Figures 1, 2, and 3 present the results obtained using the PSNR, SSIM, and ADM metrics for a dataset of 50 cover images. The PSNR results shown in Figure 1 indicate that the 2-LSB method achieves nearly constant values of approximately 44 dB for all tested images, whereas the PVD method yields lower values. It can be observed that the PSNR values of the PVD method vary depending on the characteristics of the cover image, ranging from approximately 35 dB to 43 dB. Since higher PSNR values indicate a lower difference between the cover and the stego image, it can be concluded that the 2-LSB method provides better imperceptibility.

The SSIM results presented in Figure 2 show that both methods achieve values close to 1, indicating a high degree of similarity between the cover and the stego image. However, the PVD method yields slightly higher SSIM values, suggesting better preservation of the image's structural characteristics after the embedding process.

The results obtained using the ADM metric are presented in Figure 3. According to this metric, the PVD method achieves higher values for most of the observed images. These results indicate better image quality preservation and greater similarity between the cover and stego image when the PVD method is applied.

Analyzing the obtained results, we came to the conclusion that the choice of metric affects the imperceptibility of the method. This observation can be explained by the fact that different metrics evaluate different aspects of image quality.

Figure 4 shows the BER values obtained using the 2-LSB and PVD methods as a function of the variance of the AWGN (σ^2_{AWGN}). Lower BER values indicate greater robustness of the method, while a BER value of 0 indicates accurate and complete reconstruction of the embedded secret message. The values of σ^2_{AWGN} are calculated on the pixel range from 0 to 255.

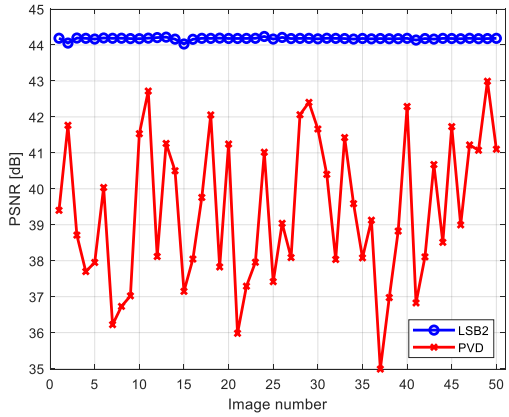


Figure 1. PSNR values for 2-LSB and PVD method

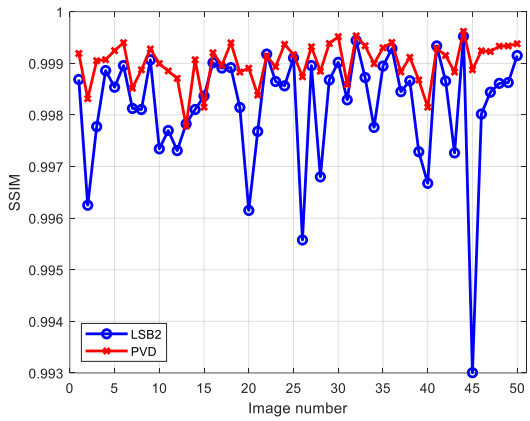


Figure 2. SSIM values for 2-LSB and PVD method

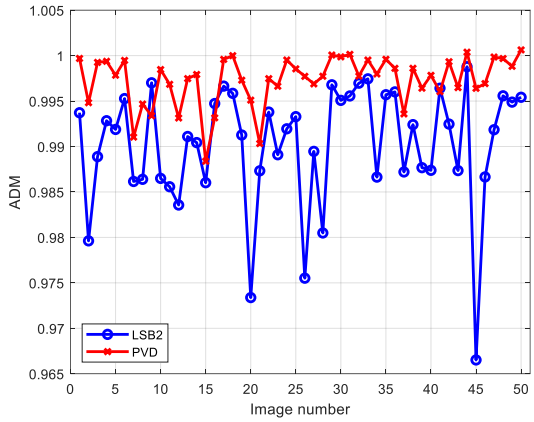


Figure 3. ADM values for 2-LSB and PVD method

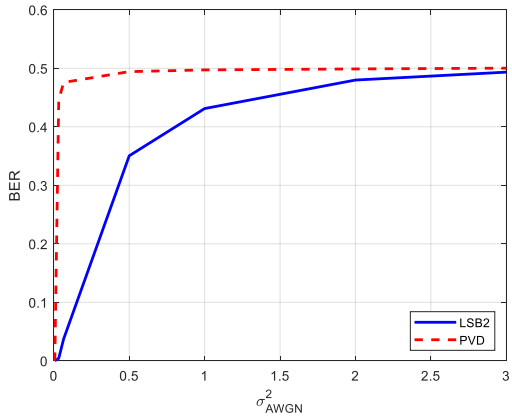
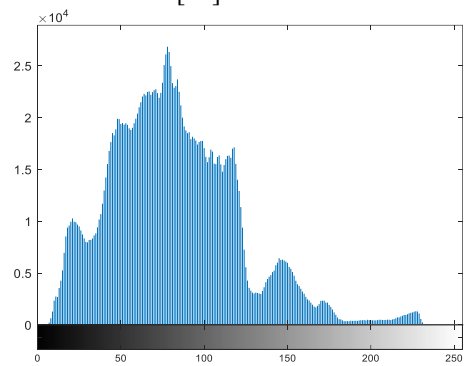
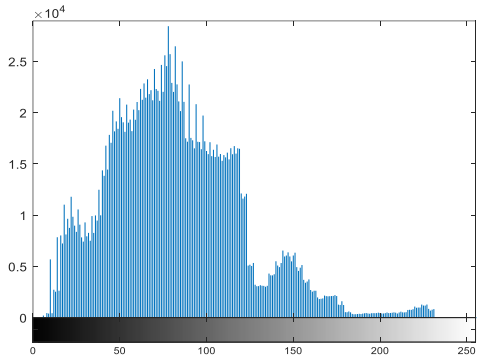


Figure 4. BER values for 2-LSB and PVD method

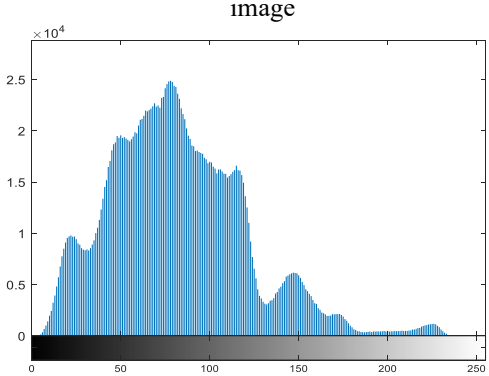
As shown in the graph, both methods achieve a BER of 0 in the absence of AWGN. It can be observed that the 2-LSB method exhibits a gradual increase in BER as the AWGN intensity increases, whereas the BER of the PVD method rises sharply even at low AWGN levels and approaches a value of 0.5. At higher AWGN levels, both methods produce BER values of approximately 0.5, making reliable reconstruction of the embedded secret message impossible. These results indicate that the 2-LSB method is more robust against AWGN than the PVD method, but both methods are highly error prone even at low noise levels. The obtained results are in accordance with previous research [14].



(a) Histogram of the cover image



(b) Histogram of a 2-LSB stego image



(c) Histogram of PVD stego image

Figure 5. Histograms of the cover and stego images

The histogram of the cover image is shown in Figure 5.a, while the histograms of the stego images obtained using the 2-LSB and PVD methods are presented in Figures 5.b and 5.c, respectively. The histograms shown correspond to one of the images from the dataset and illustrate the distribution of pixel intensity, where the horizontal axis represents the pixel intensity and the vertical axis represents the number of pixels. The histogram shown in Figure 5.b exhibits more pronounced deviations in the distribution of pixel intensity, manifested by the presence of noticeable peaks and valleys. In contrast, the histogram in Figure 5.c retains a shape that is more similar to the shape of the cover image. These observations are further confirmed by the histogram correlation and RMSE results shown in Table 1.

Correlation represents a measure of similarity between the histograms of the cover and stego images. Correlation values closer to unity indicate greater similarity of pixel intensity distributions. Based on the mean correlation values, 0.9539 for the 2-LSB method and 0.9796 for the PVD method, it can be concluded that PVD method achieves values closer to unity, and therefore the histograms are more similar to those of the cover images. RMSE is a measure of the deviation between histograms. The average RMSE values are 2613.7 for the 2-LSB method and 1753.3 for the PVD method. Smaller values indicate smaller deviations, so the obtained results confirm that the PVD method introduces smaller deviations in the histograms of the cover and stego images.

Table 2 shows the results of the assessment of the algoritham complexity of the analyzed methods. The results shown in the table indicate that embedding the secret message into the cover image using the 2-LSB method requires significantly less time than using the PVD method. The average execution time for the 2-LSB method is 0.3175 s, while for the the PVD method is 0.9491 s, meaning that the PVD method takes approximately three times longer. From the above results, it can be concluded that the 2-LSB method is more efficient in terms of the execution time, which is expected given that the PVD method involves additional calculations related to determining the range of differences and correcting pixel values.

Table 1. Results of correlation and RMSE using 2-LSB and PVD method

Methods		Metrics	
		Correlation	RMSE
2-LSB	Min	0.6859	816.68
	Mean	**0.9539**	**2613.7**
	Max	0.9931	3134.8
PVD	Min	0.8243	168.63
	Mean	**0.9796**	**1753.3**
	Max	0.9997	2769.5

Table 2. Algorithm complexity measurement results

Time [s]	Min	**Mean**	Max
2-LSB	0.2773	**0.3175**	0.3762
PVD	0.9081	**0.9491**	0.9870

Table 3. Results of measuring the capacity of the method

Capacity	Min	**Mean**	Max
2-LSB	4.147.200	**4.147.200**	4.147.200
PVD	3.114.494	**3.268.800**	3.489.559

Table 3 shows the results of measuring the capacity of the 2-LSB and the PVD method for all 50 observed images.

The capacity of the 2-LSB method is fixed and amounts to 4,147,200 bits. On the other hand, it can be observed that the capacity of the PVD method varies and depends on the characteristics of the cover image. The average capacity across all 50 images is 3,268,800 bits, which means that the PVD method allows embedding a smaller amount of data compared to the 2-LSB method. This difference in capacity can be explained by the fact that the number of bits that can be embedded in a given pixel pair using the PVD method depends on the range to which the pixel difference belongs, whereas in the 2-LSB method it is constant.

4. CONCLUSION

This paper analyzes the effectiveness of the steganography method based on pixel value differencing. The obtained results were compared with those obtained using the method based on the replacement of the two least significant bits. The analysis was conducted in the context of imperceptibility, robustness, security, algorithm complexity and capacity.

It can be concluded that neither of the evaluated methods provides the best performance across all considered criteria. Both methods exhibit specific advantages and limitations, depending on the aspect being analyzed. The obtained results show that the 2-LSB method outperforms the PVD method in terms of embedding capacity, robustness against AWGN, and computational complexity based on execution time measurements. In contrast, the imperceptibility analysis indicates that the PVD method achieves better performance according to the majority of the applied image quality metrics, relative to maximum embedding capacity. Furthermore, the PVD method demonstrates a higher level of statistical imperceptibility with respect to cover image, based on two histogram metrics. Overall, the results indicate that the choice of the embedding method depends on the specific application requirements and the desired trade-off among imperceptibility, robustness, security, embedding capacity, and computational complexity.

In future work, it is planned to analyze improved versions of the PVD method and develop own modifications of the algorithm in order to increase its efficiency. Additionally, the application of cryptography techniques prior to embedding the message in the cover image could further increase the level of security, so it would be useful to consider that possibility as well.

Acknowledgment

This research has been a part of the project No. CPME/R1/26-28 supported by the Ministry of Defence, Republic of Serbia.

References

[1] WU,W., LI,H.: *A novel method for random sequential high-capacity data hiding based on PVD and LSB*, Research Square (2023).

[2] BISHT,A., SINGLA,A., JOSHI.K.: *A review on image steganography techniques*, International Research Journal on Advanced Engineering Hub (IRJAEH), 2(7), (2024) 1986-1996.

[3] HAMMAD,O.M., SMAOUI,I., FAKHFAKH,A., HASHIN,M.M.: *An overview of image steganography techniques: Historical development, methodologies, and evaluation criteria*, Shifra, 2024, (2024), 74-87.

[4] JANA,S., BISWAS,R., DAS,D., CHATTERJEE,D., PAL,N., BASAK,D., PAL,K.: *A review paper of least significant bit (LSB) based steganography with encryprion*, International Journal of Scientific Research in Engineering and Menagement (IJSREM), 9(6) (2025) 1-9.

[5] ALSHAMRANI,H., ALAJMANI,S., ALYAMI,R.Y. SOH,B.: *Comprehensive comparison of image steganography techniques with security enhancement*, International Journal of Recent Technology and Engineering (IJRTE), 14(2), (2025) 7-19.

[6] JIN,L., LIN,J.Y., HU,S., WANG,H., WANG,P., KATSAVOUNIDIS,I., AARON,A., KUO,C.C.J.: *Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis*, In Proceedings of the IS&T International Symposium of Electronic Imaging - Image Quality System Performance XIII, San Francisco, CA, USA, (2016), Article number IQSP-222.

[7] WU,D.C., TSAI,W.H.: *A steganographic method for images using pixel-value differencing,* Pattern recognition letters, 24(9-10), (2003) 1613-1626.

[8] MAJEED,N.D., AL-ASKERY,A.J., HASAN,F.S., ABOOC,S.: *A survey on steganography and image encryption techniques*, Electrical Engineering Technical Journal, 2(1), (2025) 11-24.

[9] ALMALIKI,A.J.Q., GHAZALI,O., DIN.R., UTAMA,S.: *Comparative analysis of LSB, PVD, and EMD-based stenographic methods with hybrid optimization in digital images*, International Journal of Engineering and Techniques, 11(6), (2025) 351-357.

[10] BONDŽULIĆ,B., PAVLOVIĆ,B., PETROVIĆ,V. ANDRIĆ,M.: *Performance of peak signal-to-noise ratio quality assessment in video streaming with packet losses*, Eletronics Letters, 52(6), (2016) 454-456.

[11] WANG,Z., BOVIK,A.C., SHEIKH,H.R., SIMONCELLI,E.P.: *Image quality assessment: From error visibility to structural similarity*, IEEE Transactions on Image Processing, 13(4), (2004) 600-612.

[12] LI,S., ZHANG,F., MA,L., NGAN,K.N.: *Image quality assessment by separately evaluating detail losses and additive impairments*, IEEE Transactions on Multimedia, 13(5), (2011) 935-949.

[13] LIN,W.-B., LAI,T.-H., CHANG,K.-C.: *Statistical feature-based steganalysis for pixel-value differencing steganography*, EURASIP Journal on Advances in Signal Processing, 2021, (2021) 87.

CHAUDHARY,S., GUPTA,G., MISHRA,V.K., BALYAN,V., SONI,P.K.: *A review on image steganography techniques: Evolution from classical to adaptive methods*, Signals, 7(4), (2026), 78.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EVALUATION OF TRANSFORMER-BASED EMBEDDING MODELS FOR INFORMATION RETRIEVAL IN SERBIAN MILITARY TECHNICAL TEXTS

SLAVIMIR STOŠOVIĆ
The Academy of Applied Technical and Preschool Studies, Niš, slavimir.stosovic@akademijanis.edu.rs

NIKOLA VUKOTIĆ
The Academy of Applied Technical and Preschool Studies, Niš, nikola.vukotic@akademijanis.edu.rs

Abstract: *The effectiveness of text retrieval systems depends on the quality of underlying document representations. For morphologically rich languages such as Serbian, the choice of embedding model is particularly important due to complex inflectional patterns that affect lexical matching. This paper presents a systematic evaluation of five transformer-based embedding models for information retrieval in Serbian military technical texts. Three monolingual models pre-trained on Serbian corpora (BERTic, Jerteh-81, Jerteh-355) are compared against two multilingual sentence transformers (multilingual-e5-base, paraphrase-multilingual-MiniLM). The evaluation is conducted on a corpus of 566 articles from Vojnotehnicki glasnik (Military Technical Courier), the scientific journal of the Serbian Ministry of Defence, using 90 domain-specific queries with graded relevance judgments. Retrieval performance is measured using normalized discounted cumulative gain (nDCG), mean reciprocal rank (MRR), and precision at k. The results reveal that the BM25 lexical baseline and the best multilingual sentence transformer perform on par (nDCG@10 = 0.757 for both; MRR = 0.985 vs. 0.989), while monolingual Serbian models perform substantially worse (best: Jerteh-355, nDCG@10 = 0.235). This finding demonstrates that for domain-specific technical text with specialized vocabulary, lexical matching remains fully competitive with dense retrieval at a fraction of the computational cost, and that cross-lingual contrastive training, rather than monolingual pre-training, is what makes dense retrieval viable for Serbian. Monolingual models without contrastive training objectives are unsuitable for retrieval tasks. The benchmark provides practical guidance for the selection of retrieval components in search systems designed for defence organizations operating in the Serbian language.*

Keywords: *embedding models, information retrieval, Serbian language, transformer models, military technical documentation.*

1. INTRODUCTION

The rapid growth of digital documentation in military and defence organizations creates an increasing need for automated information retrieval (IR) systems capable of locating relevant documents in response to domain-specific queries. In the Serbian-speaking defence context, the primary publication venue for military technical research is Vojnotehnicki glasnik (Military Technical Courier), the scientific journal of the Serbian Ministry of Defence, which has published over 1,700 articles spanning topics from weapons systems and communications to logistics and military management.

Modern information retrieval increasingly relies on dense retrieval methods, where documents and queries are mapped to vector representations (embeddings) using neural language models, and relevance is computed as vector similarity [1]. The quality of these embeddings directly determines retrieval performance. For morphologically rich languages such as Serbian, which exhibits extensive inflectional paradigms (nouns decline for 7 cases in 2 numbers across 3 genders), the choice of embedding model is particularly consequential: a model that fails to capture morphological variants of the same lexeme will produce poor representations of both documents and queries.

Two families of transformer-based models are available for Serbian text: monolingual models pre-trained specifically on Serbian (or South Slavic) corpora, such as BERTic [2] and Jerteh [3], and multilingual models trained on diverse multilingual data, such as multilingual-e5 [4] and paraphrase-multilingual-MiniLM [5]. A priori, it is unclear which approach yields better embeddings for Serbian: monolingual models have deeper exposure to Serbian morphology and vocabulary, while multilingual models benefit from cross-lingual transfer learning and are explicitly trained for semantic similarity tasks.

To the best of our knowledge, no prior study has systematically evaluated embedding models for Serbian-language information retrieval in a technical domain. The present work addresses this gap by constructing a benchmark consisting of a corpus of 566 articles from Vojnotehnicki glasnik, a set of 90 domain-specific queries with graded relevance judgments, and a comparative evaluation of five transformer-based embedding models. The main contributions are:

Corresponding author: N. Vukotić;
E-mail: nikola.vukotic@akademijanis.edu.rs

DOI: 10.5937/OTEH26089S

1. A domain-specific benchmark for Serbian military technical text retrieval, including a corpus, query set, and relevance judgments.

2. A systematic comparison of monolingual versus multilingual embedding models on this benchmark.

3. Practical recommendations for embedding model selection in Serbian-language retrieval systems.

2. RELATED WORK

Dense retrieval using neural embeddings has been extensively studied for English [1, 6]. The BEIR benchmark [7] provides a standardized evaluation framework across multiple domains and retrieval tasks, but does not include Serbian or any South Slavic language.

For Serbian, BERTic [2] is a BERT-based model pre-trained on Bosnian, Croatian, Montenegrin, and Serbian web corpora, primarily designed for token-level tasks (POS tagging, NER) rather than sentence-level similarity. The Jerteh models [3] are encoder models pre-trained specifically on Serbian text, available in 81M and 355M parameter variants. Neither BERTic nor Jerteh was trained with a contrastive or sentence similarity objective, which is known to be important for producing effective retrieval embeddings [8].

Multilingual sentence transformers, such as multilingual-e5 [4] and paraphrase-multilingual-MiniLM [5], are trained on large-scale multilingual parallel data with explicit contrastive objectives. These models support 50-100+ languages and have demonstrated strong cross-lingual transfer performance on sentence similarity benchmarks [9].

Retrieval from technical documents has been studied for high-resource languages, including a dedicated technical-documents CLIR task at TREC [10], but comparable work for Serbian technical texts, military or civilian is absent from the literature. The companion paper [11] extends the present evaluation by integrating the best-performing retrieval model into a complete Retrieval-Augmented Generation system with a locally deployed LLM.

Serbian is a morphologically rich language with complex inflectional patterns (7 cases, 3 genders, 2 numbers) that affect lexical matching in retrieval systems [12]. Available NLP resources include the srLex morphological dictionary with 1.86 million word forms [13]. Lexical retrieval methods such as BM25 [14] may benefit from morphological normalization (lemmatization) as a preprocessing step, a direction we explore in future work.

3. METHODOLOGY

3.1. Corpus construction

The evaluation corpus was constructed from publicly available articles published in Vojnotehnicki glasnik between 2014 and 2025. Article metadata was retrieved via the Crossref API, yielding 1,752 records. Full-text PDFs were downloaded from the journal's website and the SCIndeks platform. After text extraction and cleaning (removal of headers, footers, references, and non-textual content), the final corpus contains 566 articles with a total of 16.6 MB of clean text.

Vojnotehnicki glasnik was selected as the evaluation source for two reasons: operational defence documentation (technical manuals, standards, and internal reports) is sensitive and cannot be used in a public evaluation, while the journal offers publicly available texts that share the terminology and style of military technical writing. This choice keeps the benchmark fully reproducible and shareable, while the findings are expected to transfer to the broader class of defence documents.

For retrieval, documents were segmented into chunks based on structural boundaries (title, abstract, body sections), producing 15,514 chunks with an average length of 1,247 characters. This chunking strategy preserves semantic coherence within each segment and reflects the granularity at which users typically seek information.

3.2 Query set and relevance judgments

A set of 90 queries was constructed to cover three categories of information needs:

- Title queries (30): queries derived from article titles, representing exact-match retrieval scenarios (e.g., "evaluation of ballistic protection materials").
- Abstract-derived queries (30): queries paraphrasing key findings from article abstracts, representing semantic similarity scenarios (e.g., "impact of temperature on composite armor performance").
- Domain-specific queries (30): general military-technical questions not tied to a specific article (e.g., "communication systems used in Serbian armed forces").

Relevance judgments were assigned on a 4-point scale: 0 (not relevant), 1 (marginally relevant), 2 (relevant), 3 (highly relevant), and constructed by pooling: for each query, the top-10 results of all six evaluated systems were merged, yielding 3,547 unique query-document pairs. Chunks of the source document for title and abstract-derived queries were assigned the highest grade by the known-item convention. Each pooled pair was independently graded by two locally deployed open-weight language models (Gemma 3 4B and Qwen 2.5 7B); a third model (Llama 3.1 8B) adjudicated disagreements by majority, which resolved 94% of pairs automatically, and the remaining 220 pairs (6%) were graded by a domain expert. Excerpts consisting mainly of reference lists were treated as at most marginally relevant. The judging protocol was calibrated on 35 expert-graded pairs, reaching within-one-grade agreement with the expert on 97% of automatically resolved pairs. An initial manual expert review of an earlier, programmatically generated label set had found systematic over-scoring of marginally related pairs (and a latent lexical bias that

could unfairly favor sparse retrieval), which motivated this protocol.

3.3 Embedding models

Five transformer-based models were evaluated (Table 1), representing the two families described above.

For base encoder models (BERTic, Jerteh), embeddings were obtained by mean-pooling the last hidden layer token representations. For sentence transformer models (e5, MiniLM), the native encoding interface was used, which applies an appropriate pooling strategy internally.

Table 1. Embedding models evaluated

Model	Type	Parameters	Languages	Embedding dim.	Size (MB)
BERTic [2]	Base encoder	111M	BCMS	768	420
Jerteh-81 [3]	Base encoder	81M	Serbian	768	310
Jerteh-355 [3]	Base encoder	355M	Serbian	1024	1,351
multilingual-e5-base [4]	Sentence transformer	278M	100+	768	1,061
paraphrase-multilingual-MiniLM [5]	Sentence transformer	118M	50+	384	449

3.4 Evaluation protocol

For each model, all 15,514 corpus chunks were encoded into dense vectors. For each of the 90 queries, the query was encoded using the same model, and the top-k most similar chunks were retrieved using cosine similarity. Retrieval performance was measured using:

- nDCG@k (normalized discounted cumulative gain): measures ranking quality accounting for graded relevance, at cutoffs k = 1, 5, 10, 20.
- MRR (mean reciprocal rank): average of the reciprocal rank of the first relevant result.
- P@k (precision at k): fraction of retrieved documents that are relevant.

All experiments were conducted on a single consumer machine (Apple M2, 16 GB RAM) without a discrete GPU; models ran on the integrated GPU via the Metal (MPS) backend.

4. RESULTS

4.1 Overall retrieval performance

Table 2 presents the main retrieval results across all 90 queries.

Table 2. Retrieval performance across all 90 queries.

Model	Type	nDCG@5	nDCG@10	MRR
BM25	Sparse	0.793	0.757	0.985
multilingual-e5-base	Dense	0.788	0.757	0.989
paraphrase-MiniLM	Dense	0.752	0.701	0.994
Jerteh-355	Dense	0.266	0.235	0.659
Jerteh-81	Dense	0.244	0.213	0.634
BERTic	Dense	0.067	0.064	0.278

The BM25 lexical baseline and multilingual-e5-base achieve statistically indistinguishable overall performance (nDCG@10 = 0.757 for both; paired permutation test, $p = 0.99$), demonstrating that keyword-based matching remains fully competitive for domain-specific technical text where queries and documents share specialized terminology. Both significantly outperform paraphrase-MiniLM (nDCG@10 = 0.701, $p < 0.01$). The monolingual Serbian models perform substantially worse, with Jerteh-355 achieving only nDCG@10 = 0.235 and BERTic scoring very low (nDCG@10 = 0.064).

Three key findings emerge. First, BM25 matches the best dense model on this benchmark while requiring no neural encoding, indicating that the high lexical overlap between technical queries and domain-specific documents makes exact term matching a strong baseline. Second, multilingual sentence transformers dramatically outperform monolingual models (3.2-fold improvement in nDCG@10 of e5-base over Jerteh-355). Third, BERTic, despite being the most widely used South Slavic language model, is unsuitable for retrieval, confirming that models not trained with a sentence-level similarity objective produce poor document representations.

4.2 Performance by query type

Table 3 presents nDCG@10 and MRR broken down by query type.

Table 3. Retrieval performance by query type (nDCG@10 / MRR).

Model	Title	Abstract-derived	Domain-specific
BM25	0.855 / 0.978	0.758 / 1.000	0.660 / 0.978
e5-base	0.825 / 0.983	0.765 / 0.983	0.681 / 1.000
MiniLM	0.751 / 0.983	0.714 / 1.000	0.636 / 1.000
Jerteh-355	0.198 / 0.574	0.303 / 0.793	0.203 / 0.611
Jerteh-81	0.163 / 0.568	0.251 / 0.680	0.226 / 0.654
BERTic	0.031 / 0.152	0.051 / 0.234	0.112 / 0.448

BM25 leads on title queries (nDCG@10: 0.855 vs. 0.825 for e5-base), where exact term overlap is strongest, while e5-base leads on abstract-derived (0.765 vs. 0.758) and domain-specific queries (0.681 vs. 0.660), where semantic matching beyond exact keywords pays off. MRR is close to its ceiling for all three strong systems across query types (0.978-1.000), so ranking quality differences are visible mainly in nDCG. All models perform worst on

domain-specific queries, which require broader topical matching beyond exact keyword overlap.

4.3 Computational efficiency

Table 4 presents the processing speed and resource requirements.

The paraphrase-multilingual-MiniLM model is by far the fastest, encoding the entire corpus in 3.2 minutes at 79.9 chunks per second - over 10 times faster than e5-base and 53 times faster than Jerteh-355. Given that MiniLM also achieves the second-best dense retrieval performance (nDCG@10 = 0.701, MRR = 0.994), it represents the best accuracy-efficiency trade-off for practical deployment.

Table 4. Computational efficiency (corpus: 15,514 chunks, Apple M2, no discrete GPU).

Model	Encoding speed (chunks/s)	Corpus encoding time	Model size (MB)
BERTic	6.7	38.9 min	420
Jerteh-81	14.0	18.5 min	310
Jerteh-355	1.5	2.8 hours	1,351
e5-base	7.4	34.7 min	1,061
MiniLM	79.9	3.2 min	449

5. DISCUSSION

The evaluation reveals two central findings: (1) BM25 lexical retrieval performs on par with the best dense model on this domain-specific benchmark at a fraction of the computational cost, and (2) among dense models, multilingual sentence transformers dramatically outperform monolingual Serbian models.

Why lexical retrieval remains competitive on technical text. The strong performance of BM25 (nDCG@10 = 0.757, on par with the best dense model) is explained by the nature of military technical documentation: articles use a highly specialized and consistent vocabulary where queries and relevant documents share many of the same technical terms. In such domains, exact keyword matching naturally achieves high precision. This finding aligns with observations in the BEIR benchmark [7], which showed that BM25 remains competitive with dense models on domain-specific corpora with specialized terminology.

Why monolingual dense models fail. The monolingual models (BERTic, Jerteh) were pre-trained with masked language modeling (MLM), which learns contextual word representations but does not optimize for sentence-level semantic similarity. The contrastive training objective used by multilingual sentence transformers directly optimizes the property needed for retrieval: semantically similar texts should have similar vector representations. This explains the 3.2-fold gap in nDCG@10 between e5-base (0.757) and Jerteh-355 (0.235). BERTic's very low retrieval scores (nDCG@10 = 0.064) confirm that widely used language models are not automatically suitable for all downstream tasks.

Cross-lingual transfer compensates for monolingual depth. The multilingual models benefit from training on parallel data across 50-100+ languages, enabling knowledge transfer from high-resource languages to Serbian. This cross-lingual alignment implicitly maps morphological variants to similar regions of the embedding space, providing robustness to Serbian's rich inflectional morphology.

Practical recommendations. For deployment in defence organizations, the choice between sparse and dense retrieval depends on the application scenario. BM25 is recommended when queries use domain-specific terminology: it requires no GPU, minimal memory, and matches the accuracy of the best dense model on this benchmark. For applications requiring semantic matching (e.g., cross-language queries, paraphrase detection), multilingual-e5-base provides the best dense alternative. Paraphrase-multilingual-MiniLM offers the best speed-accuracy trade-off for real-time applications (10x faster encoding than e5-base). A hybrid approach combining BM25 with dense retrieval [1] is likely to provide the best overall performance, as the two methods capture complementary relevance signals.

6. CONCLUSION

This paper presented the first systematic benchmark of retrieval models for Serbian military technical texts, comparing BM25 lexical retrieval against five transformer-based dense models on a corpus of 566 articles from *Vojnotehnicki glasnik* with 90 domain-specific queries. The results demonstrate that BM25 performs on par with the strongest dense model (nDCG@10 = 0.757 for both) while requiring no neural encoding, confirming the continued effectiveness of lexical matching for domain-specific technical text. Among dense models, multilingual sentence transformers (e5-base, MiniLM) achieve competitive performance and substantially outperform monolingual Serbian models (BERTic, Jerteh), with the training objective (contrastive vs. MLM) identified as the primary factor explaining this gap.

Future work will explore hybrid retrieval approaches combining BM25 with dense models to capture both lexical and semantic relevance signals, domain-adaptive fine-tuning of multilingual models on Serbian military text, and the integration of morphological normalization (lemmatization) as a preprocessing step to improve lexical matching for morphologically rich Serbian text.

References

[1] KARPUKHIN,V., OGUZ,B., MIN,S., et al.: Dense passage retrieval for open-domain question answering, Proc. EMNLP, 2020, pp. 6769-6781.

[2] LJUBEŠIĆ,N., LAUC,D.: BERTic - the transformer language model for Bosnian, Croatian, Montenegrin and Serbian, Proc. 8th BSNLP Workshop, 2021, pp. 37-42.

[3] JERTEH: Serbian language models, https://huggingface.co/jerteh, 2024.

[4] WANG,L., YANG,N., HUANG,X., et al.: Multilingual E5 text embeddings: a technical report, arXiv:2402.05672, 2024.

[5] REIMERS,N., GUREVYCH,I.: Making monolingual sentence embeddings multilingual using knowledge distillation, Proc. EMNLP, 2020, pp. 4512-4525.

[6] LIN,J., NOGUEIRA,R., YATES,A.: Pretrained transformers for text ranking: BERT and beyond, Synthesis Lectures on HLT, 14(4), 2021.

[7] THAKUR,N., REIMERS,N., RUCKTASCHEL,A., et al.: BEIR: A heterogeneous benchmark for zero-shot evaluation of IR models, Proc. NeurIPS, 2021.

[8] GAO,T., YAO,X., CHEN,D.: SimCSE: Simple contrastive learning of sentence embeddings, Proc. EMNLP, 2021, pp. 6894-6910.

[9] FENG,Y., YANG,S., LAN,J., et al.: Language-agnostic BERT sentence embedding, Proc. ACL, 2022, pp. 878-891.

[10] LAWRIE,D., MACAVANEY,S., MAYFIELD,J., et al.: Overview of the TREC 2024 NeuCLIR Track, arXiv:2509.14355, 2025.

[11] VUKOTIĆ,N., STOŠOVIĆ,S.: A retrieval-augmented generation system for Serbian military technical documentation using local language models, Proc. 12th OTEH, Tara, Serbia, 2026.

[12] ERJAVEC,T.: MULTEXT-East: morphosyntactic resources for Central and South-Eastern European languages, Language Resources and Evaluation, 46(1), 2012, pp. 131-142.

[13] LJUBEŠIĆ,N., KLUBICKA,F., AGIC,Z., JAZBEC,I.P.: New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian, Proc. LREC, 2016, pp. 4264-4270.

[14] ROBERTSON,S., ZARAGOZA,H.: The probabilistic relevance framework: BM25 and beyond, Foundations and Trends in Information Retrieval, 3(4), 2009, pp. 333-389.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A RETRIEVAL-AUGMENTED GENERATION SYSTEM FOR SERBIAN MILITARY TECHNICAL DOCUMENTATION USING LOCAL LANGUAGE MODELS

NIKOLA VUKOTIĆ
The Academy of Applied Technical and Preschool Studies, Niš, nikola.vukotic@akademijanis.edu.rs

SLAVIMIR STOŠOVIĆ
The Academy of Applied Technical and Preschool Studies, Niš, slavimir.stosovic@akademijanis.edu.rs

Abstract: *The growing volume of military technical documentation in the Serbian language creates a need for efficient search and question-answering systems that can provide accurate answers to domain-specific queries. This paper presents a Retrieval-Augmented Generation (RAG) system designed for Serbian military technical texts, built entirely on locally deployed open-source components to ensure data sovereignty and eliminate dependence on external cloud services. The system combines BM25 lexical retrieval with a locally hosted large language model (Gemma 3 4B) for answer generation grounded in retrieved document segments. The system is evaluated on a corpus of 566 publicly available articles from Vojnotehnicki glasnik (Military Technical Courier), the scientific journal of the Serbian Ministry of Defence, using 90 domain-specific queries. Retrieval performance achieves an MRR of 0.985 and nDCG@10 of 0.757, while the generation component produces domain-relevant responses for 98% of queries with an average response time of 45 seconds per query on consumer hardware without a discrete GPU. A companion study demonstrates that BM25 matches the best dense neural retrieval model on this benchmark at a fraction of the computational cost, motivating the choice of lexical retrieval as the primary retrieval component. The results demonstrate that a fully local RAG system can serve as a practical tool for knowledge management within defence organizations operating in the Serbian language.*

Keywords: *retrieval-augmented generation, Serbian language, military technical documentation, information retrieval, large language models.*

1. INTRODUCTION

Defence organizations generate and maintain large volumes of technical documentation, including standards, regulations, research articles, technical manuals, and operational reports. Locating specific information within these collections is a time-consuming task that traditionally relies on manual search or basic keyword queries. The emergence of Retrieval-Augmented Generation (RAG) systems [2] offers a promising alternative: rather than returning a list of matching documents, a RAG system retrieves relevant passages and uses a large language model (LLM) to synthesize a direct answer grounded in the source material.

However, deploying RAG systems in defence contexts presents unique challenges. First, military documentation often contains sensitive information that cannot be transmitted to external cloud services, necessitating fully local deployment [3]. Second, Serbian is a morphologically rich language with complex inflectional patterns that affect both retrieval and generation quality [4]. Third, domain-specific military terminology requires specialized evaluation to ensure that retrieved passages are genuinely relevant to the query.

This paper presents a RAG system for Serbian military technical documentation that addresses these challenges through three design decisions: (1) all components run locally on consumer hardware without internet connectivity or a discrete GPU, ensuring data sovereignty; (2) BM25 lexical retrieval is used as the primary retrieval method, motivated by our companion study [1] which demonstrated that BM25 matches the best neural dense retrieval model on this domain-specific corpus at a fraction of the computational cost; and (3) a lightweight locally deployed LLM (Gemma 3 4B) generates answers in Serbian, grounded in retrieved passages.

The system is evaluated on a corpus of 566 articles from Vojnotehnicki glasnik (Military Technical Courier), the scientific journal of the Serbian Ministry of Defence, using 90 domain-specific queries covering three categories of information needs. The main contributions are:

1. A complete RAG architecture for Serbian military technical text that operates entirely on local infrastructure without cloud dependencies.

2. An empirical evaluation demonstrating that BM25-based retrieval combined with a 4B-parameter local LLM achieves practical performance (MRR = 0.985, 98% query coverage) on consumer hardware.

3. A reproducible benchmark consisting of a publicly

Corresponding author: N. Vukotić;
E-mail: nikola.vukotic@akademijanis.edu.rs
DOI: 10.5937/OTEH26090V

available corpus, query set, and evaluation framework for Serbian-language military technical retrieval.

2. RELATED WORK

RAG was introduced by Lewis et al. [2] as a method for combining retrieval and generation to produce factually grounded outputs. The approach has since been applied to domain-specific question answering, including passage-level retrieval with generative models [5] and retrieval from large-scale token databases [6]. Gao et al. [7] provide a comprehensive survey of RAG architectures, categorizing approaches by retrieval method (sparse vs. dense vs. hybrid) and generation strategy.

For morphologically rich languages, retrieval quality depends heavily on the ability to match inflected word forms. In our companion study [1], we demonstrated that BM25 lexical retrieval matches the best dense neural model (multilingual-e5-base) on technical text in Serbian language (nDCG@10 = 0.757 for both), while monolingual models such as BERTic are unsuitable for retrieval. Equal accuracy at a fraction of the computational cost motivates our choice of BM25 as the retrieval backbone.

The deployment of LLMs on local infrastructure has become increasingly viable with the release of efficient small models such as Llama 3 [8], Gemma [9], Mistral [10], and Phi-3 [11]. Frameworks such as Ollama [12] and llama.cpp [13] enable inference on consumer CPUs without GPU hardware. For defence applications, local deployment is essential to maintain data sovereignty and comply with information security requirements [3].

Serbian presents specific challenges for NLP due to its rich inflectional morphology (nouns decline for 7 cases, 2 numbers, 3 genders), two-script system (Latin and Cyrillic), and dialectal variation (ekavian vs. ijekavian) [14]. Available language resources include the srLex morphological dictionary [15], the SETimes.SR corpus [16], and transformer models such as BERTic [17] and Jerteh [18]. In prior work, we developed POS tagging systems for Serbian [19, 20] with detailed error analysis on the MULTEXT-East tagset [21]. The present work extends this line of research to information retrieval and question answering in a specialized domain.

3. SYSTEM ARCHITECTURE

The RAG system consists of three components operating in sequence: document preprocessing, retrieval, and answer generation. The complete pipeline runs locally on consumer hardware (Apple M2, 16 GB RAM) without internet connectivity, using only the integrated GPU.

3.1 Corpus and preprocessing

The evaluation corpus consists of 566 articles published in Vojnotehnicki glasnik between 2014 and 2025. Article metadata was collected via the Crossref API, and full-text PDFs were downloaded from the journal website and SCIndeks platform. Text was extracted from PDFs and cleaned to remove headers, footers, reference lists, and non-textual content, producing 16.6 MB of clean text.

For retrieval, documents were segmented into 15,514 chunks based on structural boundaries (title, abstract, body sections), with an average chunk length of 1,247 characters. This granularity balances retrieval precision (shorter chunks are more focused) with context sufficiency (longer chunks provide more information for answer generation).

3.2 BM25 retrieval

BM25 (Best Matching 25) [22] is a probabilistic lexical retrieval model that ranks documents by term frequency, inverse document frequency, and document length normalization. We use the Okapi BM25 implementation with default parameters ($k1 = 1.5$, $b = 0.75$).

For each query, BM25 scores are computed against all 15,514 corpus chunks, and the top-k chunks ($k = 5$) are selected as context for answer generation. The choice of BM25 over dense retrieval is empirically motivated: our companion study [1] showed that BM25 achieves an MRR of 0.985 and nDCG@10 of 0.757 on this corpus, statistically indistinguishable from the best dense model (multilingual-e5-base) while requiring no neural encoding. This competitiveness stems from the high lexical overlap between technical queries and domain-specific documents containing specialized vocabulary.

The BM25 index is built in 3.1 seconds and requires negligible memory beyond the corpus text, making it suitable for deployment on resource-constrained hardware.

3.3 Answer generation with local LLM

Retrieved chunks are assembled into a context window and passed to a locally deployed LLM along with the user query. We use Gemma 3 4B [9], a 4-billion parameter language model from Google, served locally via the Ollama framework [12]. The model runs locally on an Apple M2 (integrated GPU via the Metal backend), with no discrete GPU required.

The generation prompt, shown below, instructs the model to answer based only on the provided context, to respond in Serbian (ekavian dialect), and to indicate when the context does not contain sufficient information:

"You are a knowledgeable assistant for Serbian military technical documentation. Answer the following question based ONLY on the provided context. If the context does not contain enough information, say so. Always answer in Serbian language using ekavian dialect. CONTEXT: {retrieved chunks} QUESTION: {query} ANSWER:"

Generation parameters are set to temperature = 0.3 and top-p = 0.9 to favor focused, factual responses over creative output.

The choice of Gemma 3 4B is motivated by three factors: (1) it fits in 3.3 GB of memory, enabling deployment on consumer hardware; (2) it supports Serbian text through

multilingual pre-training; and (3) it generates coherent responses typically within about 40 seconds per query, which is acceptable for non-real-time information retrieval scenarios.

4. EVALUATION

4.1 Query set

A set of 90 queries was constructed across three categories:

- Title queries (30): queries derived from article titles, representing exact-match retrieval scenarios.
- Abstract-derived queries (30): queries paraphrasing key findings from abstracts, testing semantic understanding.
- Domain-specific queries (30): general military-technical questions not tied to specific articles.

Relevance judgments were assigned on a 4-point scale (0-3) by pooling the top-10 results of six retrieval systems and grading each pooled pair with an ensemble of three locally deployed open-weight language models, with disagreements adjudicated by a domain expert; the full protocol, its calibration against 35 expert-graded pairs, and the resulting judgments are described in the companion study [1].

4.2 Retrieval performance

Table 1 presents the retrieval performance of the BM25 backbone across query types.

Table 1. Retrieval performance (BM25) established in our companion study [1], serving as the retrieval backbone for the RAG system.

Metric	All queries	Title	Abstract	Domain
MRR	0.985	0.978	1.000	0.978
nDCG@5	0.793	0.905	0.853	0.621
nDCG@10	0.757	0.855	0.758	0.660
P@5	0.918	0.947	0.887	0.920

BM25 retrieval achieves strong performance across all query types, with MRR close to its ceiling everywhere (0.978-1.000). Ranking quality is highest for title queries (nDCG@10 = 0.855), which share exact terminology with target documents, and lowest for domain-specific queries (nDCG@10 = 0.660), which require broader topical matching.

4.3 Answer generation quality

Table 2 presents the answer generation statistics.

The system generates domain-relevant responses for 98% of queries. Two queries resulted in a refusal due to insufficient context. Answers average 621 characters in length, providing concise summaries of the relevant information.

Generation time averages 44.7 ± 31.8 seconds per query (median 38.2 seconds), with the large variance reflecting differences in answer length; this is acceptable for document analysis and research tasks but would need optimization for interactive applications. The entire system (BM25 index + LLM) fits within 4 GB of memory, well within the capacity of modern consumer hardware.

Table 2. Answer generation statistics (Gemma 3 4B, no discrete GPU).

Metric	Value
Queries answered	90/90 (100%)
Non-trivial answers	88/90 (98%)
Refusals (insufficient context)	2/90 (2%)
Average answer length	621 characters
Average generation time	44.7 ± 31.8 s (median 38.2 s)
Model size	3.3 GB
Hardware	Apple M2, 16 GB RAM, no discrete GPU

4.4 Answer quality examples

Table 3 presents representative examples of system outputs across query types.

Table 3. Representative query-answer pairs.

Query type	Query (abbreviated)	Answer (abbreviated)
Title	Bidirectional stack decoding of polar codes	The algorithm uses sequential decoding from both left and right sides to reduce latency...
Abstract	Role of Serbian Army in chemical hazard management	The text describes procedures for CBRN accident response, including decontamination and evacuation...
Domain	Thermal barrier coatings for military equipment	Thermal barrier plasma spray coatings protect turbo jet engine parts from high temperatures, oxidation and corrosion...

The examples illustrate that the system generates responses that reference specific information from the retrieved context and provide domain-specific technical details. However, a systematic evaluation of answer faithfulness (e.g., through human annotation or automated metrics such as RAGAS) remains necessary to quantify the rate of hallucinated content and is left for future work.

5. DISCUSSION

The results demonstrate that a fully local RAG system, combining BM25 retrieval with a small (4B parameter) LLM, achieves practical performance for Serbian military

technical documentation. Several aspects merit discussion.

Our companion study [1] showed that BM25 matches the best dense neural model on this corpus; equal accuracy with a negligible resource footprint makes it the natural choice for the retrieval component. This finding is consistent with observations in the broader IR literature [23], where BM25 remains competitive on domain-specific corpora with specialized vocabulary. For defence applications, BM25 offers additional advantages: it requires no neural model loading, builds indexes in seconds, and operates with minimal memory.

The complete system runs on consumer hardware (Apple M2, 16 GB RAM) without a discrete GPU or internet connectivity. The total resource footprint is approximately 4 GB (3.3 GB for the LLM, 0.7 GB for the corpus and index). This demonstrates that RAG systems for specialized domains no longer require cloud infrastructure, enabling deployment in air-gapped environments where data sovereignty is paramount.

The current system has several limitations. First, the 45-second average response time, while acceptable for research tasks, is too slow for interactive applications; a discrete GPU or model quantization would address this. Second, the LLM occasionally generates responses mixing Serbian and Croatian forms, reflecting the multilingual training data; fine-tuning on Serbian-specific data would improve language consistency. Third, answer quality evaluation in this study relies on automated metrics and qualitative inspection; a systematic human evaluation with multiple annotators would provide stronger evidence of answer quality. Fourth, the system does not currently support hybrid retrieval (combining BM25 with dense models), which our companion study suggests could further improve performance.

In prior work [19, 20, 21], we developed POS tagging tools for Serbian that improve downstream NLP accuracy. Integrating lemmatization as a preprocessing step for BM25 retrieval could improve recall by matching morphological variants of query terms, representing a promising direction for future work.

6. CONCLUSION

This paper presented a Retrieval-Augmented Generation system for Serbian military technical documentation, built entirely on locally deployed open-source components. The system combines BM25 lexical retrieval (MRR = 0.985) with a locally hosted Gemma 3 4B language model for answer generation, processing queries in 45 seconds on average on consumer hardware without a discrete GPU. The complete system fits within 4 GB of memory and requires no internet connectivity, making it suitable for deployment in air-gapped defence environments. Evaluation on 90 queries against 566 articles from Vojnotehnicki glasnik demonstrates that the system produces domain-relevant responses for 98% of queries.

Future work will explore hybrid retrieval combining BM25 with dense models for improved semantic matching, GPU-accelerated inference for interactive response times, fine-tuning the LLM on Serbian military technical data, and systematic human evaluation of answer quality.

References

[1] STOŠOVIĆ,S., VUKOTIĆ,N.: Evaluation of transformer-based embedding models for information retrieval in Serbian military technical texts, Proc. 12th OTEH, Tara, Serbia, 2026.

[2] LEWIS,P., PEREZ,E., PIKTUS,A., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks, Proc. NeurIPS, 2020, pp. 9459-9474.

[3] NATO STANAG 4774: Confidentiality metadata label syntax, NATO Standardization Office, 2018.

[4] ERJAVEC,T.: MULTEXT-East: morphosyntactic resources for Central and South-Eastern European languages, Language Resources and Evaluation, 46(1), 2012, pp. 131-142.

[5] IZACARD,G., GRAVE,E.: Leveraging passage retrieval with generative models for open domain question answering, Proc. EACL, 2021, pp. 874-880.

[6] BORGEAUD,S., MENSCH,A., HOFFMANN,J., et al.: Improving language models by retrieving from trillions of tokens, Proc. ICML, 2022, pp. 2206-2240.

[7] GAO,Y., XIONG,Y., WANG,S., et al.: Retrieval-augmented generation for large language models: A survey, arXiv:2312.10997, 2024.

[8] GRATTAFIORI,A., et al.: The Llama 3 herd of models, arXiv:2407.21783, 2024.

[9] GEMMA TEAM: Gemma 3 technical report, arXiv:2503.19786, 2025.

[10] JIANG,A.Q., SABLAYROLLES,A., MENSCH,A., et al.: Mistral 7B, arXiv:2310.06825, 2023.

[11] ABDIN,M., JACOBS,S.A., AWAN,A.A., et al.: Phi-3 technical report: A highly capable language model locally on your phone, arXiv:2404.14219, 2024.

[12] OLLAMA: Run large language models locally, https://ollama.com, 2024.

[13] GERGANOV,G.: llama.cpp: Port of Facebook's LLaMA model in C/C++, https://github.com/ggml-org/llama.cpp, 2024.

[14] LJUBEŠIĆ,N., KLUBICKA,F.: {bs,hr,sr}WaC - web corpora of Bosnian, Croatian and Serbian, Proc. WaC-9, Gothenburg, Sweden, 2014, pp. 29-35.

[15] LJUBEŠIĆ,N., KLUBICKA,F., AGIC,Z., JAZBEC,I.P.: New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian, Proc. LREC, 2016, pp. 4264-4270.

[16] BATANOVIĆ,V., LJUBEŠIĆ,N., SAMARDŽIĆ,T., ERJAVEC,T.: SETimes.SR - a reference training corpus of Serbian, Proc. JT-DH, Ljubljana, 2018, pp. 11-17.

[17] LJUBEŠIĆ,N., LAUC,D.: BERTic - the transformer language model for Bosnian, Croatian, Montenegrin and Serbian, Proc. 8th BSNLP, 2021, pp. 37-42.

[18] JERTEH: Serbian language models, https://huggingface.co/jerteh, 2024.

[19] VUKOTIĆ,N., PETROVIĆ,S., STOJKOVIĆ,S.: Implementation of the Serbian language POS taggers using the NLTK library, Proc. IcETRAN/ICEST, Serbia, 2023.

[20] VUKOTIĆ,N., STOJKOVIĆ,S.: BERT-based implementation of the Serbian language POS tagger, Proc. 12th IcETRAN, Cacak, Serbia, 2025.

[21] VUKOTIĆ,N., STOJKOVIĆ,S., STOJANOVIC,J.: POS tagging and error analysis for Serbian using NLTK and the MULTEXT-East tagset, Proc. 13th IcETRAN, Serbia, 2026.

[22] ROBERTSON,S., ZARAGOZA,H.: The probabilistic relevance framework: BM25 and beyond, Foundations and Trends in Information Retrieval, 3(4), 2009, pp. 333-389.

[23] THAKUR,N., REIMERS,N., RUCKTASCHEL,A., et al.: BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models, Proc. NeurIPS, 2021.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DEEPFAKE DETECTION IN DEFENSE COMMUNICATION SYSTEMS: EVALUATION OF REAL-TIME DETECTION APPROACHES

MILAN RISTIĆ
The Academy of Applied Technical and Preschool Studies, Niš, milan.ristic@akademijanis.edu.rs

NIKOLA VUKOTIĆ
The Academy of Applied Technical and Preschool Studies, Niš, nikola.vukotic@akademijanis.edu.rs

Abstract: *The increasing sophistication of deepfake generation methods poses a direct threat to the integrity of video and audio communication channels used in defense organizations. Manipulated content can be employed for identity spoofing during video conferences, fabrication of false orders, or targeted disinformation campaigns against military personnel. This paper examines three deepfake detection architectures, two convolutional neural networks (XceptionNet and EfficientNet-B4) and a Vision Transformer (ViT-B/16), with respect to the two constraints that characterize defense communication channels: video compression and limited processing time. Detection accuracy under compression is analyzed on the basis of published results on the FaceForensics++ benchmark at its three compression levels, while the per-frame inference latency of all three architectures is measured on GPU and CPU hardware, in single and half precision, to assess their suitability for real-time deployment. The analysis shows that the reported accuracy of the selected architectures degrades significantly under heavy compression, used here as a proxy for bandwidth-limited military channels, highlighting the gap between benchmark performance and operational readiness. We discuss practical implications for integrating deepfake detection into existing defense communication infrastructure and propose a tiered detection strategy that balances computational cost with detection reliability in resource-constrained environments.*

Keywords: *deepfake detection, defense communications, real-time detection, convolutional neural networks, vision transformers, cybersecurity.*

1. INTRODUCTION

Modern defense organizations increasingly rely on digital video and audio channels for command and control, coordination between units, and communication with allied structures. The same technological progress that enables this connectivity has also produced generative models capable of synthesizing convincing fake video and audio content, commonly referred to as deepfakes [1], [2]. In a military context, the consequences of successful manipulation extend far beyond reputational damage: a fabricated video order, a spoofed participant in a command video conference, or a synthetic voice message attributed to a superior officer can directly influence operational decisions. The war in Ukraine provided an early demonstration of this threat when a manipulated video of President Zelensky calling on soldiers to lay down their arms was distributed through compromised media channels in March 2022 [3].

Research on deepfake detection has produced a range of deep learning models that achieve excellent results on academic benchmarks [1], [4]. However, published evaluations are typically conducted on high-quality video material and without constraints on processing time. Defense communication systems operate under fundamentally different conditions: tactical channels are bandwidth-limited, which forces aggressive video compression, and detection must often be performed in real time, on hardware that ranges from data-center servers to embedded devices at the tactical edge. A model that performs well on uncompressed benchmark videos but fails on heavily compressed streams, or one that is accurate but too slow for streaming analysis, is of limited operational value.

This paper examines the gap between benchmark performance and operational readiness for three representative detection architectures: XceptionNet [5], EfficientNet-B4 [6], and the Vision Transformer ViT-B/16 [7]. The contribution is threefold. First, we analyze published detection results on the FaceForensics++ benchmark [4] across its three compression levels, using compression as a proxy for the degradation present in bandwidth-limited military channels. Second, we measure the per-frame inference latency of all three architectures on GPU and CPU hardware, in single and half precision, to establish which deployment scenarios permit real-time operation. Third, on the basis of both analyses we propose a tiered detection strategy that distributes detection workload across endpoint, accelerated, and procedural layers in accordance with available computational resources.

The remainder of the paper is organized as follows. Section 2 reviews deepfake threats specific to defense communication systems. Section 3 describes the three

Corresponding author: M. Ristić;
E-mail: milan.ristic@akademijanis.edu.rs
DOI: 10.5937/OTEH26091R

evaluated architectures. Section 4 defines the evaluation methodology. Section 5 presents and discusses the results, Section 6 introduces the tiered detection strategy, and Section 7 concludes the paper.

2. DEEPFAKE THREATS TO DEFENSE COMMUNICATION SYSTEMS

Deepfake content is generated by deep generative models, most notably generative adversarial networks [8], autoencoders, and, more recently, diffusion and transformer-based architectures [2]. The wide availability of these tools has lowered the technical barrier to producing convincing synthetic media, and studies have shown that AI-generated faces are frequently perceived as more trustworthy than real ones [9], which makes human observers an unreliable last line of defense.

In defense communication systems, four threat categories are of particular concern. The first is identity spoofing in video teleconferencing. Real-time face reenactment and face-swapping techniques allow an adversary who has compromised a communication channel to impersonate a legitimate participant, gaining access to sensitive discussions or injecting false information directly into the decision-making process. The second is the fabrication of orders. A synthetic video or audio message attributed to a commanding officer exploits the inherent authority of the chain of command; under time pressure, subordinate units may act on such an order before its authenticity can be verified through independent channels. The documented use of a cloned executive voice to authorize a fraudulent transfer of 243,000 USD in 2019 [10] demonstrates that voice-based authority spoofing is already operationally viable, and the military command context differs from the corporate one only in the severity of consequences.

The third category is targeted disinformation against personnel and the public. Deepfake material tailored to local language and context can be used to demoralize troops, discredit commanders, or manufacture apparent evidence of events that never occurred, with the Zelensky video [3] as the canonical example. Once such content enters social media distribution, cascading amplification makes subsequent correction slow and only partially effective. The fourth category is the manipulation of biometric authentication. Face and voice verification systems used for access control are vulnerable to presentation and injection attacks based on synthetic samples [2], which undermines single-factor biometric authentication in security-critical facilities and systems.

A common property of all four categories is that the attack surface is the communication channel itself, and the content traversing it is compressed, resized, and re-encoded by the transport infrastructure. Any detection mechanism intended for this environment must therefore be evaluated under compression and latency constraints, not only on pristine benchmark data.

3. DETECTION ARCHITECTURES

Learning-based deepfake detection is dominated by two architectural families: convolutional neural networks and vision transformers. From each family we select representative models that are widely used in the detection literature and for which published benchmark results are available.

XceptionNet [5] is a 71-layer convolutional network built entirely from depthwise separable convolutions, which factorize spatial filtering and channel mixing into separate operations. It was adopted as the reference detector in the FaceForensics++ study [4] and has since served as the de facto baseline of the field, owing to its strong ability to capture the local, pixel-level artifacts introduced by face manipulation pipelines. With the two-class (real/fake) output layer used here in place of the ImageNet head, the model has 20.8 million parameters (22.9 million with the ImageNet head) and processes input at 299×299 pixels.

EfficientNet-B4 [6] belongs to a family of networks obtained by compound scaling of depth, width, and input resolution under a fixed computational budget. It processes input at 380×380 pixels and has 17.6 million parameters (19.3 million with the ImageNet head). EfficientNet backbones dominated the top solutions of the DeepFake Detection Challenge [11], the winner being based on EfficientNet-B7 [12]; EfficientNet-B4 and its attention-augmented variant form the backbone of the ensemble of Bonettini et al. [13], which matches or exceeds XceptionNet on FaceForensics++.

The Vision Transformer [7] departs from convolutional processing entirely: the input image is divided into fixed-size patches that are linearly embedded and processed by a stack of self-attention layers. The ViT-B/16 variant used here operates on 224×224 inputs divided into 16×16 patches and, with a two-class output layer, has 85.8 million parameters (86.6 million with the ImageNet head). Self-attention provides a global receptive field from the first layer, which allows the model to relate distant image regions, a property hypothesized to help detect the global inconsistencies of manipulated faces, complementing the local artifact sensitivity of CNNs. Transformer-based detectors have reported competitive results on FaceForensics++ across compression levels [14].

These three architectures span the relevant design space along two axes: local versus global feature extraction, and parameter count versus computational structure. As Section 5 shows, the relationship between model size, theoretical computational cost, and actual inference latency is not straightforward, which has direct consequences for deployment planning.

4. EVALUATION METHODOLOGY

4.1. Detection accuracy analysis

Detection accuracy is analyzed on the basis of published results on FaceForensics++ [4]; no detector was trained or tested by the authors. The dataset contains 1,000 original videos and their counterparts manipulated by four methods (DeepFakes, Face2Face, FaceSwap, NeuralTextures), more than 1.8 million frames in total, distributed at three compression levels: raw, c23 (H.264,

constant rate factor 23, visually near-lossless), and c40 (constant rate factor 40, strong compression with visible artifacts). The c40 level serves as a proxy for the video quality of bandwidth-limited tactical channels, not as a measurement of such links, whose encoders typically operate under rate control and are additionally affected by packet loss. Since published per-compression comparisons predominantly report accuracy, Section 5 uses accuracy as the common basis and reports AUC separately where only AUC is available.

4.2. Latency measurement

Inference latency was measured with PyTorch 2.11.0 [15] and timm 1.0.27 [16] (CUDA 12.8) on Google Colaboratory, with an NVIDIA Tesla T4 GPU (16 GB) and an Intel Xeon CPU (2 vCPUs). cuDNN autotuning (cudnn.benchmark) was enabled and the default channels-first (NCHW) memory format was used, i.e. channels_last was not enabled. Each model, instantiated with a two-class output layer, was run in inference mode with batch size 1 at its native resolution (299×299, 380×380, and 224×224 pixels), corresponding to frame-by-frame streaming processing. On the GPU, 300 forward passes were timed with CUDA events and explicit synchronization after 50 warm-up iterations; on the CPU, 30 passes were timed with a high-resolution wall-clock timer after 5 warm-up iterations, using both threads. Besides FP32 inference on both devices, FP16 GPU inference (weights and inputs converted to half precision) was evaluated, as reduced precision is the typical mode of edge accelerators; these results come from a single GPU type and are not claimed to transfer to other accelerators. We report the mean, standard deviation, and 95th percentile of latency, the maximum frame rate, and peak GPU memory in FP32. Since latency depends on architecture, input resolution, and hardware rather than on learned weights or pixel content, ImageNet-pretrained weights and random input tensors were used. Only the classifier forward pass is timed; video decoding, face detection, alignment, and preprocessing, which are common to all detectors, are excluded. The results therefore compare architectures but are not complete-pipeline latencies, in which face detection is often the most expensive stage. The real-time threshold is 33.3 ms per frame (30 FPS).

5. RESULTS AND DISCUSSION

5.1. Detection accuracy under compression

Table 1 summarizes the accuracy reported for the evaluated architectures at the three compression levels. As the sources use different protocols, the protocol is given for each entry: frame-level over all four manipulations [4], video-level over all four manipulations (Table 7, Appendix A.1 of [17]), and the Face2Face subset [14]. The values are therefore not directly comparable in absolute terms, although the compression-induced degradation is consistent across protocols; they describe the three selected architectures, not the state of the art of the field. No per-compression accuracy breakdown exists for EfficientNet-B4 (see below). Figure 1 visualizes the trend.

Table 1. Reported detection accuracy (%) on FaceForensics++ (binary real/fake classification, trained and tested at the same compression level; [17]: Table 7, Appendix A.1). EfficientNet-B4: see text after Figure 1.

Model	Protocol	raw	c23	c40	Source
XceptionNet	frame-level, all manipulations	99.26	95.73	81.00	[4]
XceptionNet	video-level, all manipulations	99.0	97.0	89.0	[17], Table 7
ViT-B/16	Face2Face subset	99.07	97.38	86.06	[14]

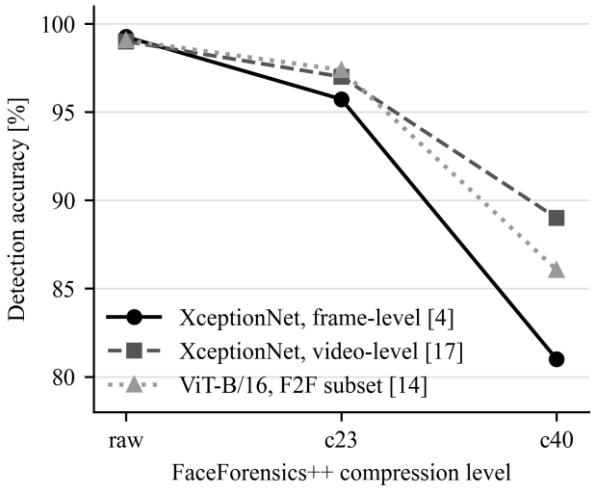


Figure 1. Reported detection accuracy as a function of FaceForensics++ compression level [4], [14], [17]

For EfficientNet-B4, Bonettini et al. [13] report at c23 an AUC of approximately 0.98 for both EfficientNet-B4 and XceptionNet in the same framework, with ensemble and attention variants giving further gains. These AUC values are not comparable with the accuracies in Table 1, and no per-compression accuracy breakdown has been published; DFDC results [11], [13], obtained on heterogeneously degraded material, suggest robustness of the same order as the CNN baseline.

Three observations follow from Table 1. First, on uncompressed material all listed models exceed 99%, which explains the optimistic perception created by headline benchmark results. Second, mild compression (c23), typical of commercial video services, costs two to four percentage points. Third, strong compression (c40) produces a qualitative change: frame-level XceptionNet accuracy drops to 81% [4], and approaches that aggregate evidence over time [17] or capture global structure [14] reach 86–89%. This range characterizes the selected architectures, not the best current detectors: compression-robust methods reach 94.2% (LipForensics) and 96.1% (RealForensics) at c40 under the same video-level protocol ([17], Table 7). The mechanism is well understood: manipulation pipelines leave high-frequency traces, and lossy compression removes precisely this content.

Operationally, the classification performance of the selected architectures is substantially reduced in the c40

regime. Overall accuracy does not show how errors divide between missed manipulations and false alarms, so the miss rate cannot be derived from Table 1; an overall error rate of 11–19%, however split, is acceptable for screening but insufficient as the sole authority for authenticating an order or a conference participant.

5.2. Inference latency

Table 2 presents our latency measurements. Figure 2 visualizes the mean per-frame latency against the 30 FPS real-time threshold.

Table 2. Measured per-frame latency of the classifier forward pass (batch size 1, native resolution): mean ± standard deviation, 95th percentile, maximum frame rate, and peak GPU memory in FP32

Model	Latency [ms]	p95 [ms]	FPS	Mem. [MB]
GPU, FP32				
XceptionNet	6.91±0.29	7.09	144.7	128
EfficientNet-B4	17.12±6.06	29.45	58.4	130
ViT-B/16	12.94±0.17	13.17	77.3	345
GPU, FP16				
XceptionNet	6.48±0.88	7.99	154.4	–
EfficientNet-B4	18.35±3.39	25.83	54.5	–
ViT-B/16	5.40±1.52	8.50	185.0	–
CPU, FP32				
XceptionNet	365.6±202.0	628.5	2.7	–
EfficientNet-B4	338.8±234.3	715.5	3.0	–
ViT-B/16	504.7±302.6	927.2	2.0	–

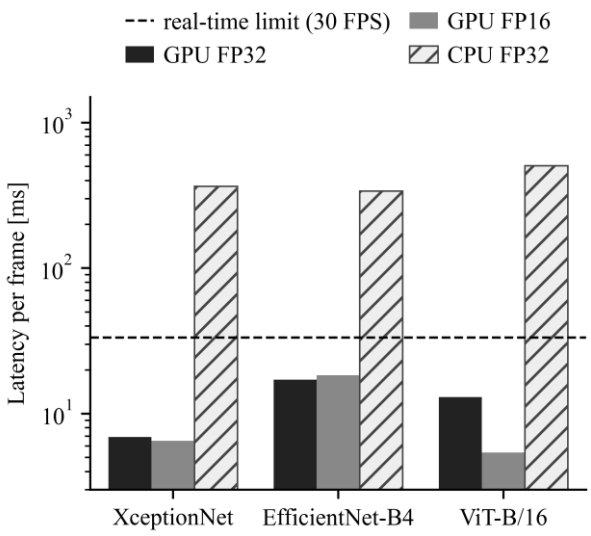


Figure 2. Per-frame inference latency on GPU (FP32/FP16) and CPU; the dashed line marks the 30 FPS real-time threshold (logarithmic scale)

On the GPU, all three architectures meet the real-time criterion in FP32, with mean latencies of 6.9–17.1 ms. EfficientNet-B4, despite the lowest theoretical cost (4.5 GMACs per frame, i.e. billions of multiply–accumulate operations as counted with fvcore, against 8.4 for XceptionNet and 16.9 for ViT-B/16), shows the highest GPU latency and by far the largest variance, with a 95th percentile of 29.45 ms, close to the 33.3 ms budget. The low arithmetic intensity of depthwise convolutions at 380×380 may contribute but cannot be the full explanation, since XceptionNet uses the same type of convolution and runs at a stable 6.9 ± 0.3 ms. Plausible further causes are the many small memory-bound kernels of the squeeze-and-excitation blocks and activations, and load from other tenants on the shared Colab GPU, which would also explain the variance; these hypotheses were not verified with a profiler. The measurement does establish that operation counts are a misleading proxy for latency. Conversely, ViT-B/16, with four times more parameters than either CNN, sustains 77 FPS in FP32, and FP16 reduces its latency by a factor of 2.4, to 5.4 ms (185 FPS), consistent with its dense matrix multiplications mapping onto tensor cores; the CNNs show marginal or no FP16 gains. On the T4 the transformer is thus the fastest model in half precision; whether this holds on other edge accelerators must be verified on the target hardware.

On the CPU, mean latencies of 340–505 ms (2–3 FPS) fall an order of magnitude short of real time, with high variance plausibly due to shared virtualized cores. Obtained with eager-mode FP32 PyTorch on two shared cores, these values are a pessimistic lower bound; even so, they support sampled screening at one frame per second, the strategy of commercial scanning services, with capacity to spare.

5.3. The operational gap

Combining both analyses quantifies the gap referenced in the title. Accuracy above 99% exists only for uncompressed material that tactical channels cannot deliver; at strong compression the selected architectures fall to 81–89%, and even compression-robust methods stay below their uncompressed level [17]. In our measurements, real-time classifier throughput required GPU acceleration, while unoptimized CPU inference is limited to sampled screening. None of the evaluated models, deployed at a single point, can therefore provide full-rate, high-confidence detection under operational conditions, which motivates the following strategy.

6. TIERED DETECTION STRATEGY

We propose organizing deepfake detection in defense communication infrastructure into three tiers that explicitly match detection workload to available resources and decision stakes. Figure 3 summarizes the proposed architecture.

Tier 1 provides continuous endpoint screening. Every endpoint, including CPU-only terminals, runs a lightweight detector on sampled frames (about 1 FPS), which is feasible without acceleration (Section 5.2). XceptionNet-class CNNs suit this tier owing to their low latency and small memory footprint (Table 2). The measured 2–3 FPS is a lower bound: optimized runtimes such as ONNX Runtime or OpenVINO with INT8 quantization can be expected to reach tens of frames per second on a workstation CPU, leaving headroom for higher sampling rates and face detection. Frame scores must be aggregated over a temporal window (e.g. averaging or majority voting over several seconds) before

an alert is raised. The purpose of this tier is triage, not authoritative classification: flagging suspicious segments for deeper analysis.

Tier 2 provides accelerated full-rate analysis. Streams flagged by Tier 1 and all high-value channels (command video conferences, order transmission) are analyzed at full frame rate on GPU-equipped nodes (data-center servers or tactical edge accelerators in FP16). The 185 FPS of ViT-B/16 on the T4 suggests that one device may serve more than one stream, but it excludes decoding, face detection, and alignment, so the number of streams per device must be determined for the complete pipeline. Combining a local-artifact CNN and a global-context transformer at this tier is intended to mitigate the single-model accuracy ceiling at strong compression, on the hypothesis, not tested here, that the two families have complementary failure modes.

The evaluated detectors are visual, whereas order fabrication and biometric spoofing (Section 2) are often audio-driven; dedicated audio detectors must therefore operate alongside them at Tiers 1 and 2, while Tier 3 covers both modalities.

Tier 3 is procedural and organizational. Because the residual error rate at operational compression levels remains too high to delegate authentication decisions to any detector, technical detection must be embedded in a zero-trust communication doctrine [18]: orders received through video or audio channels require out-of-band confirmation, participants in sensitive conferences authenticate through multiple independent factors rather than biometrics alone, and personnel are trained to treat urgency combined with channel-exclusive communication as an attack indicator. In this framework, Tiers 1 and 2 reduce the volume of manipulated content that reaches human decision points, while Tier 3 ensures that no single successful evasion translates directly into an executed false order.

The strategy is deliberately incremental: Tier 1 deploys on existing hardware immediately, Tier 2 requires targeted acceleration investment at identified choke points, and Tier 3 is a doctrinal measure whose cost is primarily organizational.

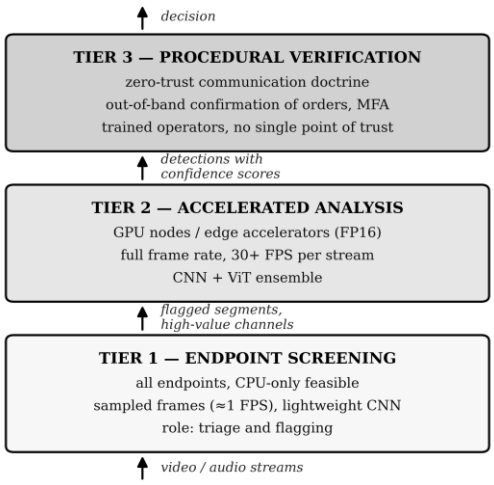


Figure 3. Tiered deepfake detection architecture for defense communication infrastructure

7. CONCLUSION

This paper examined two convolutional networks and a Vision Transformer for deepfake detection against the two constraints of defense communication: compression imposed by limited bandwidth and processing time imposed by real-time operation. Published FaceForensics++ results show the accuracy of these architectures falling from above 99% on uncompressed material to 81–89% under strong compression, while compression-robust methods reach 94–96% [17]. Our measurements of the classifier forward pass show that real-time operation is achievable on a GPU (NVIDIA T4) for all three architectures, with the Vision Transformer, contrary to expectations based on parameter count, the fastest in half precision, whereas unoptimized FP32 inference on a shared virtual CPU is limited to sampled screening. Theoretical computational cost proved an unreliable predictor of latency, underscoring the need for benchmarking on target hardware. The proposed tiered strategy assigns screening, in-depth analysis, and procedural verification to the resources capable of sustaining them. Future work will cover the complete detection pipeline, optimized CPU and embedded hardware, audio deepfake detection, and compression-robust training.

References

[1] VERDOLIVA,L.: Media Forensics and DeepFakes: An Overview, IEEE Journal of Selected Topics in Signal Processing, 14(5) (2020) 910-932.

[2] MIRSKY,Y., LEE,W.: The Creation and Detection of Deepfakes: A Survey, ACM Computing Surveys, 54(1) (2021) 1-41.

[3] Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn, NPR, March 16, 2022, https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia (accessed July 2026).

[4] RÖSSLER,A., COZZOLINO,D., VERDOLIVA,L., RIESS,C., THIES,J., NIESSNER,M.: FaceForensics++: Learning to Detect Manipulated Facial Images, Proc. IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, 2019, 1-11.

[5] CHOLLET,F.: Xception: Deep Learning with Depthwise Separable Convolutions, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017, 1251-1258.

[6] TAN,M., LE,Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Proc. International Conference on Machine Learning (ICML), Long Beach, 2019, 6105-6114.

[7] DOSOVITSKIY,A., BEYER,L., KOLESNIKOV,A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Proc. International Conference on Learning Representations (ICLR), 2021.

[8] GOODFELLOW,I., POUGET-ABADIE,J.,

MIRZA,M., et al.: Generative Adversarial Nets, Advances in Neural Information Processing Systems 27 (NeurIPS), Montreal, 2014, 2672-2680.

[9] NIGHTINGALE,S.J., FARID,H.: AI-synthesized faces are indistinguishable from real faces and more trustworthy, Proceedings of the National Academy of Sciences, 119(8) (2022) e2120481119.

[10] DAMIANI,J.: A Voice Deepfake Was Used To Scam A CEO Out Of $243,000, Forbes, September 3, 2019, https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/ (accessed July 2026).

[11] DOLHANSKY,B., BITTON,J., PFLAUM,B., LU,J., HOWES,R., WANG,M., FERRER,C.C.: The DeepFake Detection Challenge (DFDC) Dataset, arXiv preprint arXiv:2006.07397, 2020.

[12] SEFERBEKOV,S.: DFDC 1st place solution, GitHub repository, https://github.com/selimsef/dfdc_deepfake_challenge, 2020 (accessed July 2026).

[13] BONETTINI,N., CANNAS,E.D., MANDELLI,S., BONDI,L., BESTAGINI,P., TUBARO,S.: Video Face Manipulation Detection Through Ensemble of CNNs, Proc. International Conference on Pattern Recognition (ICPR), Milan, 2021, 5012-5019.

[14] WU,H., WANG,P., WANG,X., XIANG,J., GONG,R.: GGViT: Multistream Vision Transformer Network in Face2Face Facial Reenactment Detection, Proc. International Conference on Pattern Recognition (ICPR), Montreal, 2022, 2335-2341.

[15] PASZKE,A., GROSS,S., MASSA,F., et al.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, Advances in Neural Information Processing Systems 32 (NeurIPS), Vancouver, 2019, 8024-8035.

[16] WIGHTMAN,R.: PyTorch Image Models, GitHub repository, https://github.com/huggingface/pytorch-image-models, 2019.

[17] HALIASSOS,A., MIRA,R., PETRIDIS,S., PANTIC,M.: Leveraging Real Talking Faces via Self-Supervision for Robust Forgery Detection, Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, 2022, 14950-14962.

[18] ROSE,S., BORCHERT,O., MITCHELL,S., CONNELLY,S.: Zero Trust Architecture, NIST Special Publication 800-207, National Institute of Standards and Technology, Gaithersburg, 2020.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EVALUATING THE EFFECTIVENESS OF THE LOCALLY DEPLOYED GPT-OSS MODEL IN DETECTING SMB AND RDP BRUTE-FORCE ATTACKS FROM WINDOWS SECURITY LOGS

IGOR PETROVIĆ
Ministry of Defence, Belgrade, igor.petrovic@mod.gov.rs

MILOMIR JOVIĆEVIĆ
Ministry of Defence, Belgrade, milomir.jovicevic@mod.gov.rs

MARKO PREMOVIĆ
Serbian Armed Forces, Belgrade, marko.premovic@vs.rs

MLADEN VEINOVIĆ
Singidunum University, Belgrade, mveinovic@singidunum.ac.rs

IVAN TOT
University of Defence, Belgrade, ivan.tot@mod.gov.rs

Abstract: *Brute-force attacks targeting Server Message Block (SMB) and Remote Desktop Protocol (RDP) services remain among the most common techniques used to gain unauthorized access to Windows systems. Traditional detection approaches are typically based on manually defined rules and event correlation mechanisms, which may be difficult to maintain and adapt to evolving attack patterns. Recent advances in Large Language Models (LLMs) have demonstrated promising capabilities in log analysis and anomaly detection; however, their applicability to authentication-based attack detection in Windows environments remains insufficiently explored. This paper evaluates the effectiveness of a locally deployed GPT-OSS model in detecting SMB and RDP brute-force attacks from Windows Security logs. A controlled experimental environment was established consisting of a Windows host with SMB and RDP services enabled, a Kali Linux attack host executing brute-force attacks using Hydra, and an NXLog-based log collection pipeline. Authentication-related security events were preprocessed and analyzed using a context-aware approach in which the model received a sliding window of recent log entries and classified the latest event as benign or malicious. The experimental results demonstrate that the proposed approach achieves high detection performance, obtaining F1 scores of 0.999 for SMB and RDP brute-force attacks in contextual judgement. The findings indicate that combining targeted preprocessing of authentication logs with context-aware reasoning performed by a Mixture-of-Experts (MoE) LLM enables accurate identification of attack patterns without relying on handcrafted detection rules. The main contributions of this work include the development of a reproducible framework for LLM-based analysis of Windows Security logs, the creation of a labeled dataset of SMB and RDP brute-force events, and the experimental validation of GPT-OSS as an effective locally deployable solution for real-time security monitoring.*

Keywords: *RDP, SMB, brute-force attacks, GPT-OSS, LLM log analysis.*

1. INTRODUCTION

Remote administration services such as Server Message Block and Remote Desktop Protocol are widely used in enterprise and personal computing environments. The massive use of these services has opened up many opportunities for launching attacks, among which brute force attacks are very common way for gaining unauthorized access to Windows systems [1]. By repeatedly attempting different username and password combinations, attackers seek to compromise valid accounts and establish an initial foothold within a target network.

Traditional detecting approaches are primarily based on predefined rules, signatures, or threshold-based correlation mechanisms implemented within Security Information and Event Management platforms, but modern use of technology demands novel approach and automatization of detection process. In addition, the increasing volume and complexity of security logs generated by modern information systems create important challenges for security analysts responsible for monitoring and incident response. Recent advances in Large Language Models (LLMs) have demonstrated promising capabilities in understanding, analyzing, and reasoning over semi-structured and unstructured textual data [2]. Unlike conventional machine learning approaches that typically require extensive feature engineering and model training, LLMs can leverage

Corresponding author: I. Petrović;
E-mail: igor.petrovic@mod.gov.rs
DOI: 10.5937/OTEH26092P

contextual information contained within log sequences and provide interpretable explanations for their decisions. Despite these developments, the effectiveness locally deployable LLMs for detecting authentication-based attacks from Windows Security logs remains insufficiently explored, especially when it comes to mixture of experts models that more rationally use the hardware resources of the servers on which they are run.

This paper presents a proof-of-concept framework for detecting SMB and RDP brute-force attacks using a locally deployed GPT-OSS model. The experimental environment consists of a Windows host with SMB and RDP services enabled, a Kali Linux attack machine executing brute-force attacks using Hydra, and an NXLog-based log collection pipeline. Windows Security events are preprocessed and provided to the GPT-OSS model in the form of a sliding context window containing the most recent log entries. The model is then tasked with determining whether the latest event corresponds to benign activity or to an ongoing brute-force attack.

The main contributions of this work are:

(1) the development of a practical framework for LLM-based real-time analysis of Windows Security logs,

(2) the experimental evaluation of a locally deployed GPT-OSS model in detecting SMB and RDP brute-force attacks,

(3) novel dataset of attack-driven and processed log messages, generated through controlled execution of brute-force attack scenarios. and

(4) an assessment of the feasibility of employing LLMs as components of future security monitoring and threat detection systems.

2. RELATED WORKS

Brute-force attacks targeting Windows authentication services remain one of the most prevalent techniques used to gain unauthorized access to enterprise systems. In particular, SMB and RDP services are frequently exposed within organizational networks and are therefore common targets for password-guessing attacks. RDP has become one of the primary attack vectors used by adversaries to obtain initial access [4], while repeated authentication attempts against SMB services continue to represent a significant security concern in Windows environments. The large volume of authentication-related events generated during such attacks makes timely detection difficult and motivates the development of more intelligent log analysis approaches.

Recent advances in Large Language Models (LLMs) have opened new possibilities for automated security log analysis. LogPrécis demonstrated the applicability of LLMs in cybersecurity by analyzing Unix shell attacks captured in honeypot environments and generating concise yet informative interpretations of security events [5]. Similarly, Bokkena investigated the use of LLMs for predictive threat detection in IT security and reported experimental results exceeding 95% accuracy, outperforming several conventional machine learning approaches while simultaneously reducing false positive detections [6]. The growing success of LLMs in complex reasoning tasks has also encouraged their adoption for higher-level behavioral analysis, where contextual understanding is essential for identifying malicious activities [3].

The potential of LLMs for anomaly detection and semantic interpretation of system behavior has been further demonstrated by different frameworks [7,8]. Furthermore, the survey presented in [9] highlights the increasing importance of LLM-based anomaly detection and Retrieval-Augmented Generation techniques for cybersecurity applications, while also identifying a lack of research focused on real-time log analysis scenarios.

For locally deployed LLM solutions, Mixture-of-Experts (MoE) architectures have shown particular promise due to their ability to utilize hardware resources more efficiently while maintaining strong inference performance. Among these models, the locally deployed GPT-OSS model has demonstrated favorable results on systems equipped with 32 GB of VRAM, making it a suitable candidate for real-time cybersecurity applications [2]. For this reason, GPT-OSS was selected as the foundation of the detection framework evaluated in this work.

3. EXPERIMENTAL SETUP

The experimental environment was designed to evaluate the capability of a locally deployed GPT-OSS model to detect SMB and RDP brute-force attacks from Windows Security logs. The overall architecture enabled the generation of authentication-related security events, collection and preprocessing of Windows logs, and automated classification of security events using the GPT-OSS model.

3.1. Test Environment

Figure 1 illustrates the architecture of the proposed experimental framework.

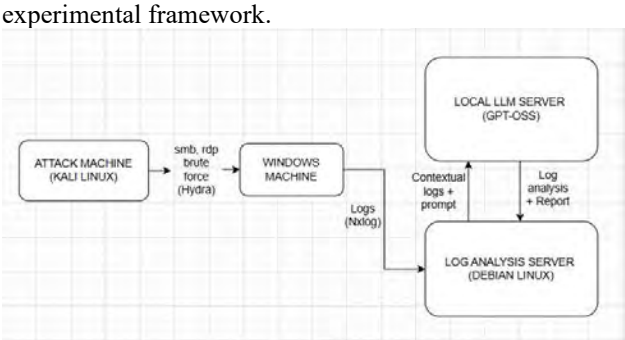


Figure 1. Experimental architecture

Figure 1 shows the architecture of the implemented framework. In order to achieve objectivity in evaluating the performance of the locally deployed GPT-OSS model, an experimental data set was created through SMB and RDP brute force attacks directed at a Windows machine. Log messages were collected and processed using the NXLOG event collection system on Windows operating

systems. The collected log messages are taken over by a specially designed application for real-time log analysis installed on a Debian Linux server and forwarded to GPT-OSS, which assigns a "malicious" or "benign" rating to the latest log message in correlation with a set of 50 previous log messages. Table 1 summarizes the main components of the experimental environment.

Table 1. Experimental environment components

Component	Description
Attacker Host	Kali Linux
Attack Tool	Hydra
Target Host	Windows system with SMB and RDP enabled
Log Collection	NXLog
AI Analysis Host	Local GPT-OSS deployment
Monitored Services	SMB and RDP

3.2. Attack Scenarios

In order to confirm the effectiveness of the locally deployed LLM GPT-OSS in assessing the maliciousness of log messages received from a Windows machine, it is necessary to establish a system in which it would be precisely known whether a log message is malicious or not. For this reason, 2 brute force attack scenarios were created that are directed at services opened on a Windows machine (SMB and RDP) with pre-prepared wordlists containing the real credentials of the machine. By launching the attack, every log message related to activity directed at the SMB and RDP services would be malicious in nature, which allows an objective assessment of the effectiveness of the observed framework. The attacks were implemented using the Hydra tool on a Kali Linux attack machine, and the attack flow is demonstrated in Figure 2.

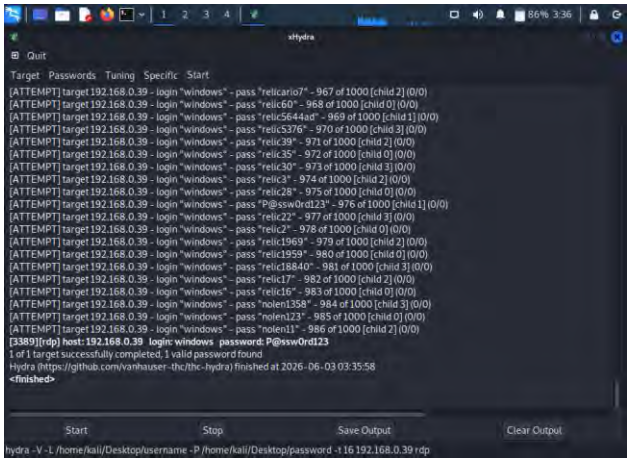


Figure 2. Generating log dataset by RDP brute-force attack using Hydra tool

3.3. Log Collection and Context Generation

Windows Security logs generated during the experiment were collected using NXLog and transformed into a structured representation containing the most relevant authentication-related attributes. Examples of extracted fields included Event ID, timestamp, username, source IP address, logon type, and authentication status.

Due to the simplicity of processing, the NXlog configuration allows for the collection of only log messages related to the SMB and RDP services, which are processed in such a way that each event is recorded as a single line in the log message, as presented:

2026-06-01 23:46:55 DESKTOP-8KVG4F0 An account failed to log on. Subject: Security ID: S-1-0-0 Account Name: - Account Domain: - Logon ID: 0x0 Logon Type: 3 Account For Which Logon Failed: Security ID: S-1-0-0 Account Name: admin Account Domain: Failure Information: Failure Reason: Unknown user name or bad password. Status: 0xC000006D Sub Status: 0xC0000064 Process Information: Caller Process ID: 0x0 Caller Process Name: - Network Information: Workstation Name: \\192.168.0.46 Source Network Address: 192.168.0.46 Source Port: 63342 Detailed Authentication Information: Logon Process: NtLmSsp Authentication Package: NTLM Transited Services: - Package Name (NTLM only): - Key Length: 0

Instead of analyzing individual events in isolation, the proposed framework used a sliding context window approach, which has proven effective in previous research using the same locally deployed model [2]. For each classification request, the GPT-OSS model received the most recent 50 log entries, including the event under investigation. This design was intended to provide temporal and contextual information that could help the model recognize repetitive authentication failures characteristic of brute-force attacks. The GPT-OSS model was instructed to analyze the supplied log sequence and determine whether the latest event represented normal activity, an SMB brute-force attack, or an RDP brute-force attack. The generated classification was subsequently compared with the ground-truth label associated with the experimental scenario.

4. GPT-OSS-BASED DETECTION APPROACH

The proposed detection framework performs real-time analysis of Windows Security logs using a locally deployed GPT-OSS model. After collection and preprocessing by NXLog, each log event is forwarded to an analysis pipeline that evaluates both the individual event and its surrounding context. To provide temporal awareness, a sliding window containing the most recent security events is maintained in memory and continuously updated as new logs arrive. For every newly received log entry, system evaluates the same log together with a context window consisting of previously observed events. This approach enables the model to identify behavioral patterns that are characteristic of brute-force attacks, such

as repeated authentication failures originating from the same source host or targeting the same user account.

The GPT-OSS model receives the log data through a structured prompt that instructs it to classify the observed activity as either benign or malicious. When malicious behavior is detected, the model is additionally required to identify the attack category and provide a short explanation supporting its decision. The generated output is subsequently stored together with the original log entry and metadata for further evaluation.

To support future analysis and reproducibility, all processed events and corresponding model decisions are archived in a vector database and exported to a structured dataset. This design enables direct comparison between classifications obtained from single-event analysis and those produced using contextual information, allowing the effectiveness of context-aware LLM-based detection to be systematically evaluated.

5. RESULTS AND DISCUSSION

As the technical feasibility of the system was presented in the previous chapter, it was necessary to confirm the efficiency of GPT-OSS in log analysis. The metric used to evaluate model performance, as well as statistical analysis, is the F1 score. The results important for calculating the F1 score (for RDP and SMB brute-force attacks) are given in Table 2.

Table 2. Confusion matrix results for single-log and contextual log analysis of logs

Attack type	GPT-OSS 20B single log judgement				GPT-OSS 20B contextual log judgement			
	TP	FP	TN	FN	TP	FP	TN	FN
RDP brute-force	975	0	0	17	991	0	0	1
SMB brute-force	761	0	0	217	976	0	0	2

The presented results represent the evaluation of the GPT-OSS 20B model listed in the table in classifying log messages as malicious or benign for two specific attack scenarios - RDP and SMB brute force attacks on Windows 10 machine. All log messages generated as a system response to attacks were considered malicious. By using parameters given in Table 2, Precision, Recall and F1 score are calculated and presented in Table 3.

Table 3. Confusion matrix results for single-log and contextual log analysis of logs

Attack type	GPT-OSS 20B single log judgement			GPT-OSS 20B contextual log judgement		
	Precision	Recall	F1	Precision	Recall	F1
RDP brute-force	1	0.983	0.991	1	0.999	0.999
SMB brute-force	1	0.778	0.875	1	0.998	0.999

Given results show that GPT-OSS 20B achieves very high accuracy in detecting both RDP and SMB brute-force attacks, with very high F1 score in all experiments, indicating a complete absence of false positive detections. However, analysis of individual logs shows a significantly lower recall, especially for SMB brute-force attacks (0.778), suggesting that the model misses a large number of malicious events when it lacks additional context. By introducing context from 50 previous logs, the performance improves drastically, with Recall reaching 0.999 for RDP and 0.998 for SMB brute-force attacks. Consequently, the F1 measure increases from 0.991 to 0.999 for RDP and from 0.875 to 0.999 for SMB, confirming that contextual information plays a key role in correctly recognizing attack patterns. The results indicate that such a system has significant potential for application as an additional layer of protection within security operations centers and incident detection and response systems. It is particularly suitable for continuous analysis of large volumes of logs and early detection of attack patterns through event correlation, which can improve the overall effectiveness of cyber protection and reduce the likelihood of missing malicious activities.

5. CONCLUSION

This paper presented an approach for applying a locally deployed GPT-OSS large language model to the detection of SMB and RDP brute-force attacks using Windows Security logs. The motivation for this research originates from the continued prevalence of authentication-based attacks against critical Windows services and the limitations of traditional rule-based detection mechanisms when dealing with large volumes of security events. To address these challenges, a proof-of-concept framework was developed that combines Windows log collection, NXLog-based preprocessing, context-aware event correlation, and LLM-driven classification.

The experimental evaluation was conducted in a controlled environment consisting of a Windows target host with SMB and RDP services enabled and a Kali Linux attack host executing brute-force attacks using Hydra. A dedicated dataset of Windows Security events was generated and labeled based on the performed attack scenarios. To improve the relevance of the information provided to the model, the preprocessing stage focused on authentication-related events and attributes associated with the monitored services, reducing noise while preserving security-critical context.

The obtained results demonstrate that the GPT-OSS model is highly effective in identifying brute-force activity from Windows Security logs. The model achieved an F1 score of 0.999 for RDP brute-force detection and SMB brute-force detection, indicating excellent classification performance across both attack scenarios in contextual judgement. These results confirm that a context-aware analysis approach enables the model to recognize repetitive authentication patterns and distinguish malicious activity from legitimate user behavior with a high degree of accuracy.

An additional contribution of this work is the demonstration that Mixture-of-Experts (MoE) architectures can be effectively deployed on commodity server hardware equipped with 32 GB of VRAM while

still delivering high detection performance. The GPT-OSS model was selected based on its previously demonstrated efficiency under these hardware constraints, and the results obtained in this study further validate its suitability for real-time cybersecurity applications. The proposed framework demonstrates that locally deployed open-weight language models can serve as practical components of modern security monitoring systems without requiring extensive feature engineering or manually crafted detection rules.

References

[1] CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY: Weak Security Controls and Practices Routinely Exploited for Initial Access, CISA, 2022.

[2] PETROVIĆ, I., VEINOVIĆ, M., ILIĆ, S., JOVIĆEVIĆ, M.: *Framework for Semantic Threat Detection in Docker Container Environments with Local MoE LLMs*, Electronics, 15(8):1664 (2026), DOI: https://doi.org/10.3390/electronics15081664

[3] CUI, M., et al.: AGLHunter: Automated Threat Hunting Using In-Context Learning-Enhanced LLM, 2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Compiegne, France, 2025, pp. 1812–1819, DOI: https://doi.org/10.1109/CSCWD64889.2025.11033364.

[4] VITLA, S.: *Unsecured Remote Desktop Protocol (RDP) Access: A Gateway for Ransomware Attacks and Corporate Extortion*, Journal of Computer Science and Technology Studies, 6(2) (2024) 150–165. DOI: 10.32996/jcsts.2024.6.2.17.

[5] BOFFA, M., DRAGO, I., MELLIA, M., VASSIO, L., GIORDANO, D., VALENTIM, R., BEN HOUIDI, Z.: LogPrécis: Unleashing Language Models for Automated Malicious Log Analysis: Précis: A Concise Summary of Essential Points, Statements, or Facts, Computers & Security, 141 (2024) 103805, DOI: https://doi.org/10.1016/j.cose.2024.103805

[6] BOKKENA, B.: *Enhancing IT Security with LLM-Powered Predictive Threat Intelligence*, Proceedings of the 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, 751–756, DOI: 10.1109/ICOSEC61587.2024.10722712

[7] ZHANG, H., SHAO, S., LI, S., ZHONG, Z.: SmartGuard: Leveraging Large Language Models for Network Attack Detection Through Audit Log Analysis and Summarization, arXiv:2506.16981, 2025.

[8] LI, C., ZHU, Z., ZHANG, X.: RedChronos: A Large Language Model-Based Log Analysis System for Insider Threat Detection in Enterprises, arXiv:2503.02702v2, 2025.

[9] AKHTAT, S., KHAN, S., PARKINSON, S.: *LLM-Based Event Log Analysis Techniques: A Survey*, arXiv:2502.00677v1, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

POSITION BIAS IN MULTILINGUAL LANGUAGE MODELS APPLIED TO SERBIAN

MILOMIR JOVIĆEVIĆ
Ministry of Defence, Belgrade, milomir.jovicevic@mod.gov.rs

SLAVIŠA ILIĆ
University of Defence, Belgrade, ilic.slavisa@gmail.com

MILAN GNJATOVIĆ
University of Criminal Investigation and Police Studies, Belgrade, milan.gnjatovic@kpu.edu.rs

KRISTINA ŽIVANOVIĆ
School of Electrical Engineering, University of Belgrade, Belgrade, zk245024p@student.etf.bg.ac.rs

Abstract: *Despite rapid progress in evaluating large language models, the Lost in the Middle phenomenon — a persistent weakness of Transformer-based models whereby information near the middle of the context window exerts less influence on generation — remains unresolved. This paper examines the phenomenon in Serbian, among the first such studies outside English, evaluating three recent open-weight small language models on a filtered multi-document QA dataset. Each query contains a question and ten documents, only one of which holds the answer; its position is shifted from first to tenth place to measure the resulting change in accuracy. A dedicated evaluation procedure was developed to account for the morphological complexity of Serbian, combining string-based matching with a final LLM-based judgment stage (Claude Haiku) for cases obscured by grammatical case variation. Within this dataset and context length, a Cochran's Q test revealed no statistically significant accuracy degradation for any of the three models as the gold document approached the middle of the context. Larger differences were instead observed between the models: Gemma-4-26B-A4B (92.9%) and Qwen3.6-35B-A3B (91.9%) could not be distinguished statistically ($p = 0.151$), while GPT-OSS-20B (87.3%) was reliably weaker than both ($p < 0.0001$), suggesting that model choice may play a more important role than gold-document position.*

Keywords: *position bias, lost in the middle, information retrieval, language-specific bias, open-weight LLMs*

1. INTRODUCTION

Large language models largely represent the most advanced technology in the world of natural language processing. Although very widespread in everyday use, the Transformer architecture, as the basis of most contemporary language models, faces many problems that have not been completely overcome to date. Most of these shortcomings are related to the context window, a component that in many ways negatively affects the accuracy of the model. One of the most well-known disadvantages of language models is referred to as the Lost in the Middle phenomenon. It assumes that the performance of the model is higher when important information is located at the beginning or at the end of the input context, while on the other hand, the ability of the model to find relevant information drops significantly when position of relevant information tends to the middle of the context window [1].

The research on this phenomenon is of great importance primarily due to a significant number of systems based on language models and designed for tasks such as Retrieval-Augmented Generation, information retrieval and multi-document QA. These systems have in common that finding an answer is directly related to the described phenomenon, which significantly affects their adoption in production environments. So far, most of the research dealing with this problem is focused on the English language, while only a few experiments considered other languages. Accordingly, this research considers the Lost in the Middle phenomenon in the context of the Serbian language.

Since there is no narrowly specialized language model for the Serbian, in this research we consider three open-weight general-purpose multilingual language models: Qwen3.6-35B-A3B, Gemma-4-26B-A4B and GPT-OSS-20B. For none of these models has already been examined whether the Lost in the Middle phenomenon has the same correlation with all the languages or whether the position bias reflects differently for different languages. Since a multilingual model shares all of its parameters across languages, language-specific behavior arises from the composition of the training data and from the tokenizer, and can therefore induce specific phenomena that manifest differently for different models and languages. The Lost in the Middle phenomenon is not exclusively related to open-weight models, but is also present in large language models such as GPT-4 and Claude 3 Opus [2,3].

Corresponding author: M. Jovićević;
E-mail: milomir.jovicevic@mod.gov.rs
DOI: 10.5937/OTEH26093J

However, this research is aimed at examining the Lost in the Middle phenomenon in language models applied to Serbian, which is a highly inflective and morphologically rich language.

The main contributions of this work are:

(1) This research is one of the first studies on the Lost in the Middle phenomenon in language models applied on Serbian.

(2) An evaluation procedure was developed that accounts for the morphological complexity of the Serbian language.

(3) A comparative evaluation of three current small open-weight models was conducted, showing that Gemma-4-26B-A4B and Qwen3.6-35B-A3B perform comparably in extractive question answering in Serbian, while GPT-OSS-20B is reliably weaker.

2. RELATED WORKS

The Lost in the Middle phenomenon was first described in [1], where the performance of language models, plotted against the position of the relevant information, forms a characteristic "U-curve". Later work proposed ways to mitigate it without fine-tuning, such as Multi-scale Positional Encoding [4] and rescaling a single dimension of the hidden states at inference time, which improves results by up to 15% [5]. Other work examined its causes, showing that residual connections significantly affect cumulative attention and thereby contribute to the U-curve [6]. Menschikov et al. showed that position effects differ across languages and pointed out the lack of evaluation for languages with fewer speakers [3]. Among broader long-context benchmarks, RULER showed that models which solve vanilla "Needle In A Haystack" tests still lose accuracy as the context grows [7], and the bilingual LongBench showed that Chinese is similarly affected when the answer must be found among many documents [8].

The single-gold-document design used in this work goes back to [1], where each question is accompanied by several documents, only one of which contains the answer. Later benchmarks extended multi-document question answering in other directions: the "Leave No Document Behind" benchmark, for example, is built so that every document is relevant to the final answer, and only one of its four task types (spotlight locating) places the evidence in a single document [9]. In this work, the distractors are taken from the same topic as the gold document, which is expected to make the task more demanding than when the documents come from different topics.

3. DATASET

At the moment of writing this paper, there was no publicly available dataset aimed at the research and evaluation of the Lost in the Middle phenomenon for Serbian language. Generally speaking, the Serbian language lacks a sufficient number of data corpora that will contribute to the development of language models for this language. Therefore a significantly modified version of the SQuAD-sr-md part of the dataset [10] was used. Originally, it was defined as a QA dataset in the Serbian language, suitable for training or fine-tuning models for extracting data from text. The dataset was created by translating texts from foreign languages using the adapted Translate-Align-Retrieve method, while also manual revision of the entire translation process was performed.
The dataset originally consists of 7484 questions and 1739 contexts (documents) in which the answer to them is found. The documents vary in length, but for each question it has been determined which document contains the necessary information for the answer.

Table 1. Dataset preview for first topic

Doc.	Question	Answer
1	Koliko akcija postoji u imovini Arsenala?	62.217
2	Koliki procenat akcija u Arsenalu je Kroenke posedovao do 2011. godine?	62.89
3	U kom filmu je prikazan FK Arsenal?	misterija Arsenalovog stadiona
4	Ko je pobedio Arsenal 2011. godine u Liga kupu?	Birmingem Sitija
5	U kom takmičenju je Arsenal prvi ušao u finale?	UEFA Lige šampiona
6	Koja je razlika u plavoj boji gostujuće sportske opreme Arsenala posle 1983.godine?	tamnijom nijansom
7	Koji brend je bio dobavljač sportske opreme pre 1994. godine?	Adidasa
8	Iz kojih zajednica dolaze pristalice Arsenala?	različitih društvenih klasa
9	Koji trener je krajem devedesetih godina prošlog veka doneo uspeh Arsenalu?	Arsena Vengera
10	Koji bivši igrač Arsenala je postao trener 1986. godine?	Džordža Grejema

To adopt this dataset for the purpose of this work, the filtering procedure was designed to ensure that the answer to each question appears in exactly one document. It was applied in four steps. First, questions were removed if their answer appeared in any other document belonging to the same topic (case sensitivity was not used), which reduced the number of questions from 7484 to 5960. Second, to account for the rich morphology of Serbian, questions were additionally removed if the stem of any word from the answer occurred in another document of the same topic; this step can only make the filtering stricter. Third, duplicate questions were eliminated, i.e., a small number of repeated questions whose answers differed only in minor variations; after the second and third steps, 5381 questions in 1650 documents remained. Finally, the documents were balanced in length:

documents whose word counts fell outside predefined lower (121) and upper (188) bounds were discarded, so that all documents in a query are of comparable length. Of the 48 topics, only the fifteen that retained at least ten documents within these bounds were kept; from each of them, ten documents were selected at random, and for each document one question was selected at random among the questions that survived filtering. This yields the final set of 150 questions and documents (fifteen topics with ten documents each).

This design ensures a consistent context length across queries: within a topic, all queries contain the same ten documents and differ only in their order, so their length is identical, while differences between topics are negligible. Furthermore, the topics span a wide range of domains (including sports, notable individuals, emotional wellbeing, music, the internet, history, and politics) so the evaluation does not depend on any narrowly specialized subject matter.

Table 1 shows the first topic of the dataset: each of its ten documents is accompanied by one question and the reference answer, which is a verbatim span of that document.

4. EXPERIMENTAL SETUP

The main concepts of the experiment are based on testing the Lost in the Middle phenomenon for open weight models that are deployed in a local infrastructure physically separated from the Internet. The goal is to investigate and evaluate the model's ability to find the answer to a given question and investigate whether the location of the document containing the answer directly affects the model's performance, while using modest models in the controlled environment.

4.1. Language Models

For the purposes of this work, we used one of the most current open-weight models of large companies whose architecture is based on Mixture-of-Experts technology and which can be deployed on consumer hardware that has 32 GB of VRAM. Three models were selected:

(1) Qwen3.6-35B-A3B, developed by Alibaba Cloud, released on April 15, 2026.

(2) Gemma-4-26B-A4B, developed by Google, released on April 2, 2026.

(3) GPT-OSS-20B, developed by OpenAI, released on August 5, 2025.

All three models were run using the llama.cpp inference framework, version 9859 (commit 4fc4ec554). Gemma-4-26B-A4B and Qwen3.6-35B-A3B were used with Q8_0 quantization; since the Q8_0 weights do not fit entirely within 32 GB of VRAM, the portion of the weights that exceeded the available VRAM was offloaded to system memory, which affects inference speed but not the model's output. GPT-OSS-20B was used as the full model in the MXFP4 format in which its weights were released, without additional quantization. The context window allocated on the server was 100,000 tokens. For Qwen3.6-35B-A3B, the temperature was set to 1.0, top-p to 0.95, and top-k to 20, while thinking mode was enabled. For Gemma-4-26B-A4B, the temperature was set to 1.0, top-p to 0.95, and top-k to 64, while thinking mode was enabled. For GPT-OSS-20B, the temperature was set to 1.0, top-p to 1.0, and top-k was disabled, while the reasoning effort was set to medium. These decoding parameters were selected according to the recommendations of the respective model developers for general-purpose tasks. The maximum output length was not limited, so generation ended at the model's end-of-sequence token. A fixed seed (42) was sent with every request. If a response did not contain the “<odgovor>” tag, or the request failed at the network level, the same request was repeated, up to three attempts in total, and the response of the last attempt was kept. All system and user prompts were written in Serbian using the Latin script.

These three models represent the latest generation models of the world's largest companies that contribute to the development of multilingual open-weight models, in the absence of language models specialized for the Serbian language.

4.2. Methodology

Each query contains all ten documents of one topic followed by the question, and exactly one of the documents, the gold document, contains the answer. Every document of a topic takes the role of the gold document in turn, and in ten consecutive queries it is placed at each position from the first to the tenth, while the nine distractors keep the same relative order. With fifteen topics of ten documents each, this gives 15 × 10 × 10 = 1500 queries per model, 150 for each gold position.

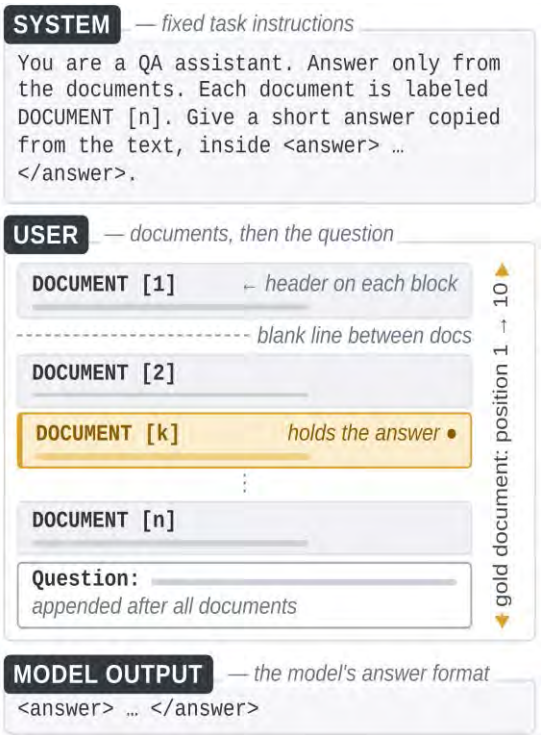


Figure 1. Structure of the prompt, with the gold document changing position across queries.

Unlike [1], which used contexts of ten, twenty, and thirty

documents and three gold positions (the first, fifth, and tenth) in its ten-document setting, only ten documents are used here, but the gold document is placed at all ten positions. In [1], accuracy was highest when the relevant document was at the beginning or the end of the context and lowest in the middle, giving the characteristic "U-curve"; comparing the edge positions with the middle ones is therefore the main aim of the design. Figure 1 shows the structure of a query. The system part instructs the model not to answer in its own words, but to return the part of the document that contains the answer inside the <odgovor> tag. This reflects the extractive nature of the dataset, in which every reference answer is a verbatim span of its document, and also makes it possible to check whether the model follows the instruction. For readability, the markers in Figure 1 are shown in English; the actual prompts use the Serbian equivalents (DOKUMENT [n], Pitanje:, <odgovor>).

5. RESULTS AND DISCUSSION

As already mentioned, the original dataset gives an answer to each question that represents a text sequence extracted from the document to which it belongs and without any change of word form. If we take into account the nature of the Serbian language, it is clear that it is not possible to use only the "exact match" method to evaluate the results of the model because the language models gave answers of different lengths, even though they all had the same instructions.

Because Serbian is a morphologically different language from English, a model can give a completely correct answer that still differs from the ground truth answer only in grammatical form — for example, a name may appear in a different case ending depending on its role in the sentence. Judging answers with a plain, character-by-character comparison would therefore unfairly mark many correct answers as wrong. To address this, each model's answers were evaluated using three checks, applied in sequence, where each approach's resulting curve is plotted and shown on figures 2, 3 and 4. First, we checked for an exact match between the model's answer and the reference answer. Second, for answers that did not match exactly, we normalized both texts (lowercasing them, removing quotation marks and trailing punctuation, and collapsing extra spaces) and compared them again, which accounts for purely cosmetic differences. Third, for answers still unmatched after the first two checks, we evaluated the remaining pairs using an LLM-as-a-judge approach: the judge model (Claude Haiku 4.5) was given the question, the reference answer, and the model's answer, together with a fixed, explicit set of criteria - grammatical and case-ending differences, synonyms, and minor differences in phrasing were to be treated as correct, while a different, incorrect, or unrelated answer was to be treated as incorrect - and asked to judge whether the two carry the same information. This step was introduced because even a permissive character-level comparison cannot recognize an answer that is phrased in an entirely different way while still conveying the correct information, which becomes increasingly relevant for longer or more paraphrased model answers. The same explicit criteria were applied uniformly to every answer, regardless of which model produced it or where the gold document had been placed in the context, and the judge was given no information about the experimental setup itself; this step was also only applied to the subset of answers left unresolved by the first two checks, rather than being used to re-evaluate every answer. Taken together, this keeps the evaluation consistent across all models and positions and confines the more permissive judgment to genuinely ambiguous cases. The judge was used on 26 September 2026 with its default decoding settings. In total, 779 answers of Gemma-4-26B-A4B, 819 of Qwen3.6-35B-A3B, and 983 of GPT-OSS-20B (2,581 overall) reached this stage, of which 673, 698, and 792 were accepted. To validate this stage, a native speaker of Serbian independently judged a random sample of the judge's decisions, stratified by model (50 per model, 150 in total), including both accepted and rejected answers. The human judgments agreed with the judge in all 150 cases (100%; 95% Wilson confidence interval 97.5–100%), and in all 50 cases for each model (95% Wilson interval 92.9–100% per model).

Because the judge never learns where the gold document was placed, the position analysis was also repeated on the outcomes of the first two stages alone, which the judge does not affect. The position effect was again not significant for any model (Cochran's Q: Gemma-4-26B-A4B $Q = 2.88$, $p = 0.969$; Qwen3.6-35B-A3B $Q = 4.44$, $p = 0.880$; GPT-OSS-20B $Q = 10.86$, $p = 0.286$), so the conclusion about position does not depend on the judge.

In the next three subsections the results per used LLM are provided. To evaluate results, three matching criteria explained earlier in this section, are shown for each evaluated LLM.

5.1. Gemma-4-26B-A4B results

Gemma-4-26B-A4B reaches the highest final accuracy of the three models (92.9% after all three evaluation stages) and shows a largely flat curve across all ten gold-document positions, ranging only from about 91% to 95% (though not significantly higher than Qwen3.6-35B-A3B, see Section 6). There is no visible dip when the gold document sits in the middle of the context, which is the pattern the original Lost in the Middle study reported for many models. As shown in Figure 2, Gemma-4-26B-A4B maintains a consistently high and stable accuracy across all gold document positions, with the 95% Wilson confidence intervals overlapping substantially at every position and no pronounced drop near the middle of the context window. A Cochran's Q test across the ten positions finds no statistically significant position effect ($Q = 5.18$, $p = 0.818$), and McNemar tests contrasting the first and the tenth position against the fifth position (the design used by Liu et al. [1]) likewise find no significant differences (92.7% and 93.3% vs. 91.3%; Holm-corrected $p = 0.750$ in both cases). The differences between the edge positions and the fifth position are +1.3 percentage points (95% CI −2.3 to +5.3) and +2.0 points (95% CI −1.3 to +5.8), so a drop toward the middle larger than about 6 points is unlikely, whereas a smaller one cannot be excluded; the position-to-position fluctuations visible in the curve therefore cannot be distinguished from sampling noise at this sample size (n=150 per position).

In 1499 of 1500 queries (99.9%), the model returned a correctly formatted <odgovor> tag on the first attempt, and only a single query remained without it. In 84.5% of cases, the extracted answer appeared verbatim in the gold document. The answers were generally short (median of three words, comparable to the reference answers), with only 7.3% exceeding ten words despite the instruction to return only a few words.

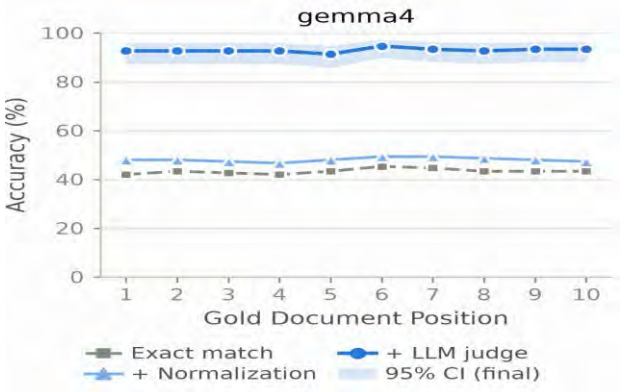


Figure 2. Gemma-4-26B-A4B accuracy by gold document position after exact matching, normalization, and LLM judge (final); shaded: 95% Wilson CI of final accuracy (n=150 per position).

5.2. Qwen3.6-35B-A3B results

Qwen3.6-35B-A3B performs close to Gemma-4-26B-A4B in absolute terms (91.9% at the final stage) and, like Gemma-4-26B-A4B, shows a largely flat curve across positions, staying within a narrow 89%–94% band with no clear dip in the middle. Its raw exact-match accuracy (42.6%) is noticeably lower than its final accuracy, and the largest jump in the cascade comes from the LLM-as-a-judge stage rather than from normalization alone, indicating that most of its apparent errors are surface-level or paraphrastic mismatches rather than genuine failures to locate the correct information. As shown in Figure 3, Qwen3.6-35B-A3B exhibits accuracy comparable to Gemma-4-26B-A4B, remaining largely stable across gold document positions and likewise showing no strong Lost in the Middle effect. A Cochran's Q test across the ten positions comes closer to conventional significance than for the other two models but still does not reach it ($Q = 15.48$, $p = 0.079$), and McNemar tests contrasting the first and the tenth position against the fifth position likewise find no significant differences (93.3% and 91.3% vs. 91.3%; Holm-corrected $p = 0.906$ and $p = 1.000$; differences +2.0 points, 95% CI −1.8 to +6.2, and 0.0 points, 95% CI −3.3 to +3.3), so the dips visible in the curve cannot be distinguished from sampling noise at this sample size. The model returned a well-formed tag on the first attempt in all 1500 queries (100%). In 79.5% of cases, the extracted answer appeared verbatim in the gold document. It also produced the shortest answers of the three models with only 5.6% exceeding ten words.

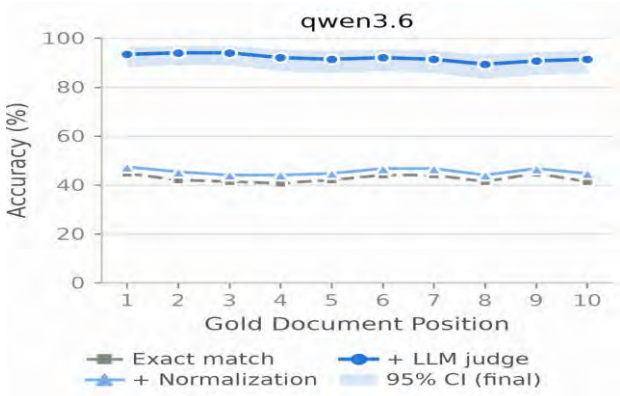


Figure 3. Qwen3.6-35B-A3B accuracy by gold document position after exact matching, normalization, and LLM judge (final); shaded: 95% Wilson CI of final accuracy (n=150 per position).

5.3. GPT-OSS-20B results

GPT-OSS-20B produces the lowest overall accuracy of the three models (87.3% at the final stage), swinging between roughly 85% and 89% across positions. The lowest point falls at position 5 and the highest at position 1, but this difference is within the range of sampling noise.

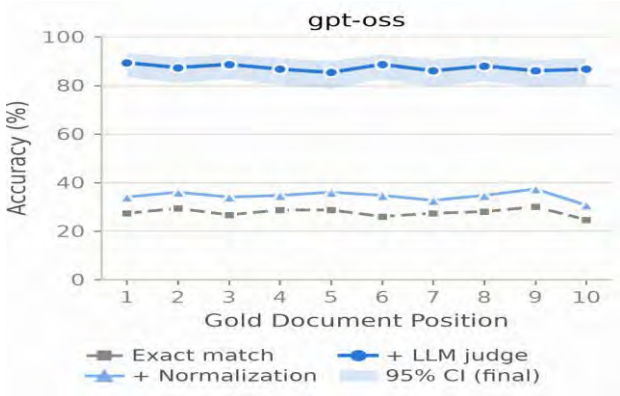


Figure 4. GPT-OSS-20B accuracy by gold document position after exact matching, normalization, and LLM judge (final); shaded: 95% Wilson CI of final accuracy (n=150 per position).

As shown in Figure 4, GPT-OSS-20B achieves the lowest overall accuracy among the tested models; however, its position-to-position swing (about 4 percentage points) is no larger than Qwen3.6-35B-A3B's (about 5 percentage points), and a Cochran's Q test across the ten positions finds no statistically significant position effect ($Q = 6.45$, $p = 0.694$) — if anything, a weaker signal than for Qwen3.6-35B-A3B. McNemar tests contrasting the first and the tenth position against the fifth position likewise show no significant differences (89.3% and 86.7% vs. 85.3%; Holm-corrected $p = 0.420$ and $p = 0.727$; differences +4.0 points, 95% CI −1.3 to +9.6, and +1.3 points, 95% CI −2.7 to +5.5). Taken together, these results suggest that GPT-OSS-20B's lower overall accuracy reflects a general weakness at extracting the correct answer rather than a heightened sensitivity to where that answer appears in the context.

A correctly formatted tag was obtained on the first attempt in 1498 of 1500 queries (99.9%), while the remaining two required a second attempt. The extracted answer appeared verbatim in the gold document in 64.8% of cases, noticeably less often than for the other two models. It also produced the longest answers (12.0% exceeded ten words), suggesting a weaker adherence to the instruction to return a short excerpt taken directly from the text.

6. CONCLUSION

This paper examined the Lost in the Middle phenomenon in the context of the Serbian language, using a filtered Multi-document QA dataset in which the position of the gold document was varied across ten positions. With awareness of the rich morphology of the Serbian language, a dedicated three-stage evaluation procedure was introduced, in which a large language model (Claude Haiku) acted as an automatic judge in the final stage. The obtained results, based on three current open-weight small language models, did not reveal a statistically significant accuracy degradation when the gold document was positioned near the middle of the context (Cochran's Q, all $p > 0.05$). The 95% confidence intervals of the edge-to-fifth-position differences reach at most about 6 percentage points for Gemma-4-26B-A4B and Qwen3.6-35B-A3B and 10 for GPT-OSS-20B, so a drop of the size Liu et al. [1] reported for GPT-3.5-Turbo in a comparable ten-document setting (about 16 points) is well outside what the data allow, whereas a drop of a few points could have gone unnoticed. This finding is specific to this dataset and these three models and suggests that the Lost in the Middle effect may not be equally manifested across all models and languages; whether this reflects the models, the length and composition of the context, or the Serbian language remains an open question for future work. By contrast, larger differences were observed between the models: GPT-OSS-20B performed reliably worse than both Gemma-4-26B-A4B (Holm-corrected $p = 1.3 \times 10^{-11}$) and Qwen3.6-35B-A3B ($p = 6.8 \times 10^{-9}$), whereas the difference between the latter two did not reach significance ($p = 0.151$), suggesting that model choice may have had a greater impact on retrieval accuracy than the position of the relevant information. The proposed evaluation methodology proved effective in correctly identifying answers despite grammatical case variation confirming that language-specific characteristics must be taken into account when evaluating large language models. For extractive question answering over ten documents in Serbian, in the configurations tested here, Gemma-4-26B-A4B and Qwen3.6-35B-A3B are equally suitable choices, whereas GPT-OSS-20B is less suitable.

References

[1] LIU, N.F., LIN, K., HEWITT, J., PARANJAPE, A., BEVILACQUA, M., PETRONI, F., LIANG, P.: Lost in the Middle: How Language Models Use Long Contexts, Transactions of the Association for Computational Linguistics 12, pp. 157-173, 2024.

[2] DSOUZA, A., GLAZE, C., SHIN, C., SALA, F.: Evaluating Language Model Context Windows: A 'Working Memory' Test and Inference-time Correction, arXiv:2407.03651, 2024.

[3] MENSCHIKOV, M., KHARITONOV, A., KOTYGA, M., PORVATOV, V., ZHUKOVSKAYA, A., KAGRAMANYAN, D., SHVETSOV, E., BURNAEV, E.: Beyond Early-Token Bias: Model-Specific and Language-Specific Position Effects in Multilingual LLMs, arXiv:2505.16134, 2025.

[4] ZHANG, Z., CHEN, R., LIU, S., YAO, Z., RUWASE, O., CHEN, B., WU, X., WANG, Z.: Found in the Middle: How Language Models Use Long Contexts Better via Plug-and-Play Positional Encoding, Advances in Neural Information Processing Systems 37 (NeurIPS 2024), Vancouver, Canada, pp. 60755–60775, 2024.

[5] YU, Y., JIANG, H., LUO, X., WU, Q., LIN, C.Y., LI, D., YANG, Y., HUANG, Y., QIU, L.: Mitigate Position Bias in Large Language Models via Scaling a Single Dimension, Findings of the Association for Computational Linguistics: ACL 2025, Vienna, Austria, pp. 6092-6111, 2025.

[6] HERASIMCHYK, H., LABRYGA, R., PRUSINA, T., LAUE, S.: A Structural Theory of Position Bias in Transformers, arXiv:2602.16837, 2026.

[7] HSIEH, C.-P., SUN, S., KRIMAN, S., ACHARYA, S., REKESH, D., JIA, F., GINSBURG, B.: RULER: What's the Real Context Size of Your Long-Context Language Models?, First Conference on Language Modeling (COLM 2024), 2024.

[8] BAI, Y., LV, X., ZHANG, J., LYU, H., TANG, J., HUANG, Z., DU, Z., LIU, X., ZENG, A., HOU, L., DONG, Y., TANG, J., LI, J.: LongBench: A Bilingual, Multitask Benchmark for Long Context Understanding, Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Bangkok, Thailand, pp. 3119–3137, 2024.

[9] WANG, M., CHEN, L., CHENG, F., LIAO, S., ZHANG, X., WU, B., YU, H., XU, N., ZHANG, L., LUO, R., LI, Y., YANG, M., HUANG, F., LI, Y.: Leave No Document Behind: Benchmarking Long-Context LLMs with Extended Multi-Doc QA, Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, Florida, USA, pp. 5627–5646, 2024.

[10] RAĐENOVIĆ, J., KITANOVIĆ, O., STANKOVIĆ, R., ŠKORIĆ, M.: Development of Serbian QA Datasets through Prompt-Based Generation and Human Validation (te-sla/QuestionAnswering), Hugging Face, https://huggingface.co/datasets/te-sla/QuestionAnswering, 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTEGRATING ARTIFICIAL INTELLIGENCE AND REVERSE ENGINEERING INTO CONSTRUCTIVE LEARNING FOR CYBERSECURITY EDUCATION IN INDUSTRY 4.0

EDIS MEKIĆ
State University of Novi Pazar, Novi Pazar, ediemceach3@gmail.com

ANA MATOVIĆ
Military Quality Control, Belgrade, ana.matovic4@gmail.com

MARIJA MATOVIĆ
Electrodistribution, Kraljevo, marija.matovic@es.rs

Abstract: *This paper will considered that the strongest instructional model is not "AI replaces analysis", but evidence-first reverse engineering, AI-assisted interpretation, human verification, secure remediation, and reflective ethical review. In the study below, a 12-week AI+RE intervention outperformed a conventional control on knowledge post-test, practical labs, retention, problem-solving quality, and ethical reasoning. The effects are intentionally plausible rather than dramatic: large enough to matter instructionally, but not so large as to imply that tools eliminate the need for human expertise.*

Keywords:*artificial intelligence (AI), reverse engineering (RE), constructive learning, cybersecurity education, software-intensive operations*

1. INTRODUCTION

Industry 4.0 combines connected production systems, software-intensive operations, and cyber-physical integration, which raises the educational value of secure software analysis, systems thinking, and workforce-aligned cybersecurity practice. In that setting, reverse engineering is not a niche specialty: it is a core method for understanding opaque software, diagnosing abnormal behavior, supporting vulnerability discovery, and building security judgment. Constructive alignment is therefore a strong pedagogical fit, because it explicitly links higher-order learning outcomes, authentic hands-on activities, and assessment. [1]

The research base also suggests a dual conclusion about large language models. LLMs can improve code explanation, tutoring, and feedback in computing education, especially when embedded in structured workflows; however, unguided access can reduce engagement for some learners, and code-focused models still generate insecure or misleading output often enough that secure-development controls and responsible-use policies are mandatory. [2]

Based on that evidence, the strongest instructional model is not "AI replaces analysis", but evidence-first reverse engineering, AI-assisted interpretation, human verification, secure remediation, and reflective ethical review. In the study below, a 12-week AI+RE intervention outperformed a conventional control on knowledge post-test, practical labs, retention, problem-solving quality, and ethical reasoning. The effects are intentionally plausible rather than dramatic: large enough to matter instructionally, but not so large as to imply that tools eliminate the need for human expertise.

The conceptual foundation of this research is a three-way synthesis. First, Industry 4.0 emphasizes digitally integrated, software-dependent production and service environments. Second, reverse-engineering research frames RE as a disciplined path from opaque artifact to actionable understanding. Third, constructive alignment and guided hands-on cybersecurity exercises show that students learn best when authentic analytical tasks are scaffolded rather than either fully scripted or completely open-ended. [3]

The semester shape assumed here is a 12-week applied course, with a one-semester intervention and a 4-week delayed retention check. That duration is adapted from the uploaded course outline, which describes a 12-week, six-credit applied computing course with prerequisite software/programming background; this report reorients that template from CMS development to cybersecurity education.

The assumed tool stack is intentionally conservative and classroom-safe: decompilation/disassembly with Ghidra; debugging and trace-based analysis with standard debuggers and syscall tracing; vulnerability vocabulary via CWE and ATT&CK; secure-development framing via SSDF; AI-risk framing via AI RMF; and training targets such as Juice Shop or DVWA that are explicitly intended for education in controlled environments [4].

Table 1. Assumed tool stack

Assumption	Value used in this report
Cohort size	72 students, randomized 1:1
Delivery model	Blended lecture + supervised lab
Duration	12 teaching weeks + 4-week retention test
Student profile	Upper-division computing/cybersecurity students with prior programming

Corresponding author: E. Mekić;
E-mail: ediemceach3@gmail.com
DOI: 10.5937/OTEH26094M

Assumption	Value used in this report
Intervention	Reverse engineering + LLM-assisted code analysis + secure coding
Comparison arm	Conventional security labs without integrated LLM workflow
LLM access	Institutionally approved interface with logging and policy controls
Lab infrastructure	Isolated VMs/containers, no public targets, sanitized artifacts only

These starting points are analytically useful because they are large enough for plausible section-level randomization, but small enough to resemble a real applied university course rather than a MOOC-scale experiment. They also align with current guidance to map educational outcomes to workforce-relevant cybersecurity knowledge and skills [5].

2. STUDY DESIGN

The study design is a mixed-method, parallel-group, randomized classroom experiment. Students are assigned to either a control condition or an AI+RE intervention. Both groups receive the same learning objectives, contact hours, instructor team, and grading weight. The difference is instructional workflow: the control arm uses conventional readings, demos, and lab writeups, while the intervention arm adds structured reverse-engineering notebooks, prompt-governed LLM support, AI critique tasks, and mandatory verification of AI-suggested explanations and remediations [6].

Table 2. Study design

Component	
Participants	72 students; Control n = 36, AI+RE n = 36
Learning activities	Reverse-engineering labs, LLM-assisted code analysis, secure coding remediation tasks
Primary knowledge measure	Pre/post cybersecurity analysis test, 0–100
Performance measures	Reverse-engineering lab score, code-analysis task score, secure-coding remediation score, practical composite
Higher-order measures	Problem-solving rubric (1–5), ethical reasoning rubric (1–5)
Retention measure	Delayed test at 4 weeks, 0–100
Qualitative data	Reflection memos, lab annotations, two focus groups, instructor observations
Statistical methods	Baseline equivalence tests; group difference in gain scores; endline group comparisons; Cohen's d for effect size; thematic coding for qualitative data
Fidelity controls	Same instructor, same capstone brief, common rubric, supervised labs, restricted tool policy

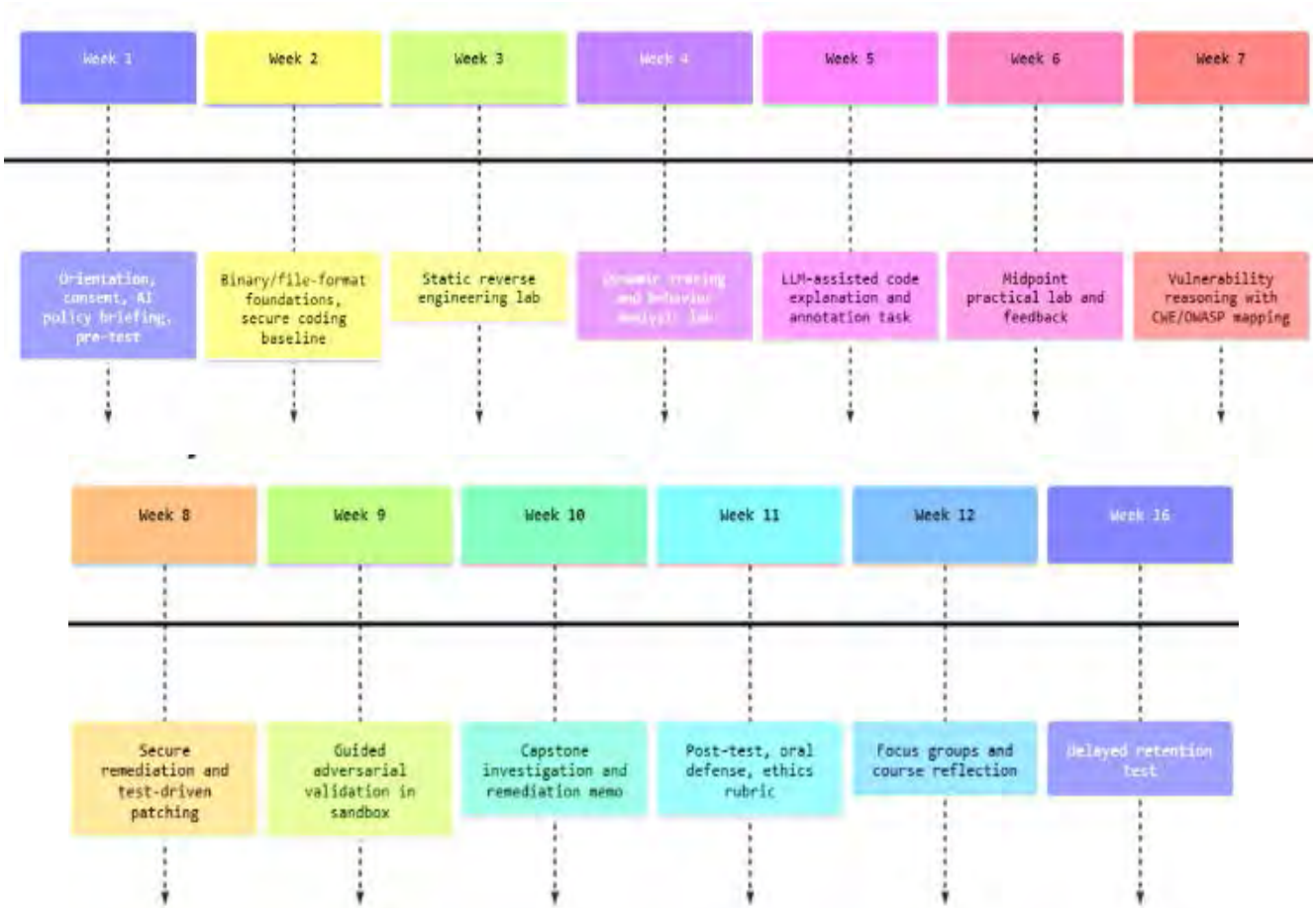


Figure 1. Design choices

A plausible weekly sequencing is to start with static structure, move to dynamic traces, then add LLM support only after students can independently produce baseline evidence. That ordering reduces the chance that students treat the model as an oracle and is consistent with findings that LLM benefits are strongest when they augment, rather than replace, student reasoning.

3. RESEARCH RESULTS

Here are empirical results from implemented study

Quantitative results

Table 3. Quantitative results

Metric	Control M(SD)	AI+RE M(SD)	p-value	Effect size d
Knowledge pre-test (0– 100)	47.8 (10.4)	48.6 (10.7)	.749	0.08
Knowledge post-test (0–100)	58.6 (11.2)	67.6 (10.9)	.001	0.81
Knowledge gain (post – pre)	10.8 (7.6)	18.6 (8.2)	< .001	0.99
Retention test at 4 weeks (0–100)	55.4 (10.3)	63.0 (10.0)	.003	0.75
Ethical reasoning rubric (1–5)	3.40 (0.50)	3.75 (0.50)	.004	0.70

Practical performance metric	Control M (SD)	AI+RE M (SD)	p-value	Effect size d
Reverse-engineering lab average (0–100)	69.0 (10.7)	76.0 (9.8)	.005	0.68
Code-analysis task (0–100)	66.5 (11.0)	78.1 (10.2)	< .001	1.09
Secure-coding remediation task (0–100)	68.2 (10.4)	77.4 (9.6)	< .001	0.92
Practical composite (0–100)	67.9 (9.8)	76.2 (9.1)	< .001	0.88
Problem-solving rubric (1–5)	2.90 (0.60)	3.35 (0.60)	.002	0.75

The largest lift appears on code explanation and remediation, not on baseline factual recall. That mirrors the current literature: LLM-assisted support tends to be most useful on feedback, explanation, and stuck-point recovery, while raw code generation still requires verification because security and engagement risks remain material.

The results also imply that AI+RE does not simply boost output speed; it improves delayed retention and ethical reasoning when the model is embedded in a supervised verification loop. That is consistent with constructive alignment and with hands-on cybersecurity work showing that analysis skills improve when exercises are designed around meaningful evidence, guided inquiry, and explicit reflection on how and why tools are used.

Qualitative findings

Table 4. Quantitative results

Theme	Simulated interpretation
Faster recovery from confusion	Students reported that LLMs helped them get unstuck on unfamiliar APIs, assembly fragments, or trace output faster than forum-style searching.
Better articulation of reasoning	Lab reports in the intervention arm contained clearer causal explanations of “what the program is doing” and “why this behavior is suspicious.”
Improved remediation quality	Students more often connected observed weaknesses to concrete mitigations and validation steps.
Trust calibration problems	A recurring complaint was that the model could sound certain while being partially wrong, especially on low-level behavior.
Ethical boundary awareness	Students said the explicit “authorized target + evidence + justification” workflow made legal and ethical boundaries feel more concrete.
Uneven dependence risk	A minority of students over-relied on the assistant early and needed instructor intervention to return to evidence-first practice.

These qualitative themes fit the current evidence base. LLMs can improve tutoring and performance for some learners, but they can also reduce engagement or encourage shallow dependence if policies, prompts, and assessment do not force verification and explanation.

Proposed Curriculum and Assessment

The most defensible instructional model is a constructive loop in which students first gather evidence from software artifacts, then use AI to accelerate interpretation, but must validate every claim through tooling, tests, or replicated observations.

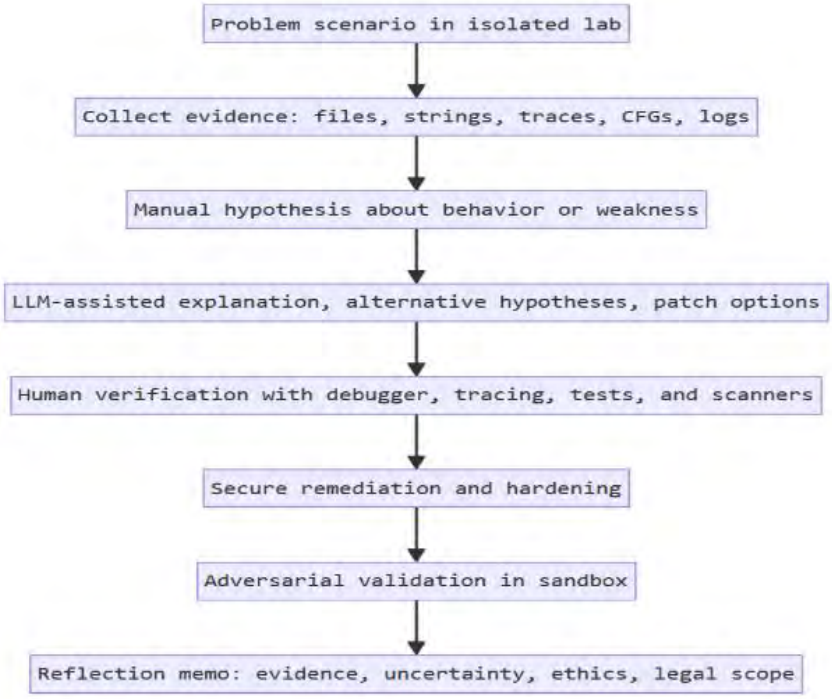


Figure 2. Instructional model (constructive loop)

That model is aligned with constructive alignment, workforce-oriented cybersecurity framing, and secure-development practice. It also matches classroom evidence that students need a middle ground between rote procedural labs and entirely unguided challenge solving.

Table 5. Constructive alignment

Curriculum block	Learning outcomes	Sample lab exercise	Assessment evidence
Foundations of secure systems and AI use	Distinguish static vs. dynamic evidence; explain where AI can and cannot be trusted	Compare a source-level flaw explanation written manually vs. with LLM assistance	Short concept quiz; AI-use policy check
Static reverse engineering	Recover program structure, identify suspicious imports/calls, annotate control flow	Analyze a toy binary or obfuscated utility in Ghidra and explain likely purpose	RE notebook; annotation quality
Dynamic behavior analysis	Validate or falsify static hypotheses with runtime evidence	Trace a trojanized utility and isolate anomalous file/network behavior	Trace lab score; oral defense
LLM-assisted code analysis	Use the model to summarize, critique, and refine code understanding while citing evidence	Map observed weakness to CWE and propose a secure patch plan	Code-analysis task; prompt log
Secure remediation and validation	Patch vulnerabilities, write tests, and confirm the fix	Repair injection/access-control flaws in a sandboxed training app	Secure-coding task; test suite
Capstone investigation	Produce an end-to-end analytical memo with ethical/legal justification	Investigate a benign simulated artifact, explain risk, and recommend mitigation	Capstone report; ethics rubric

Training targets should be deliberately educational rather than operational. Practice environments such as Juice Shop and DVWA exist specifically to support classroom-safe learning, while CWE and ATT&CK provide standardized ways to describe weaknesses and adversarial behavior.

Proposed assessment rubric

Table 6. Proposed assessment rubric

Criterion	What strong performance looks like	Weight
Evidence collection	Uses the right artifacts and tools; separates observed facts from guesses	20%
Technical explanation	Explains program behavior or weakness coherently and at the right abstraction level	20%
Vulnerability reasoning	Correctly maps issues to weakness/security concepts and prioritizes risk	20%
Secure remediation	Produces a sound patch and explains why it reduces risk without breaking functionality	20%
Validation and reflection	Verifies outcomes empirically and discusses uncertainty, ethics, and scope limits	20%

This rubric intentionally rewards reasoning quality and validation, not just "getting the right answer." That is important because LLM-era assessments must distinguish genuine understanding from polished but unverified output.

4. CONCLUSION

Recommended baseline governance is to align secure development with NIST's SSDF and AI use with its AI RMF; to use training targets and priority weakness categories from OWASP; and to structure threat and weakness language around taxonomies such as ATT&CK and CWE. In practice, that means every lab should specify the authorized target, expected evidence sources, allowed AI uses, and required validation steps before students begin [8-10].

For tooling, a defensible classroom baseline is National Security Agency's Ghidra for binary analysis; debuggers and syscall tracing for dynamic evidence; static analyzers and unit tests for secure remediation; and an institutionally governed LLM interface. If a copilot-style assistant is used, it should come with explicit responsible-use guidance, privacy controls, and a requirement that students submit prompt logs or AI-use disclosures [10-12].

For academic integrity, the key move is to assess process artifacts rather than only finished answers. Require individualized or randomized targets, evidence notebooks, short oral defenses, and a brief "what the model got right/wrong" reflection. That recommendation is supported by work showing both the educational upside of guided LLM use and the risk that generic chat access can lower engagement or encourage shallow task completion. Randomized reverse-engineering challenges are especially useful because they reduce answer sharing while preserving common learning objectives [13].

For safety and ethics, labs should run only in isolated VMs or containers with snapshot rollback, no live malware, no public targets, no institutional secrets in prompts, and no offensive deployment outside the classroom sandbox. In the United States, good-faith security-research exemptions are limited to lawful contexts and are not safe harbors against other laws. In the EU, observation, study, testing, and some decompilation are permitted only within specific legal conditions, especially interoperability and lawful use [14-19]. The course therefore needs an explicit rule: students may analyze only software and systems they own or are authorized to use for the lab's stated educational purpose.

The practical implication is straightforward: the most credible AI+RE curriculum is not the one with the most automation, but the one with the strongest verification culture [20-22]. Students should leave the course able to explain what they observed, what the model suggested, what they independently verified, what remained uncertain, and why their actions stayed within ethical and legal scope.

References

[1] Mekić, E. S., Jovanović, M. N., Kuk, K. V., Prlinčević, B. P., & Savić, A. M. (2024). Enhancing educational efficiency: Generative AI chatbots and DevOps in Education 4.0. *Computer Applications in Engineering Education*, *32*(6), e22804.

[2] Purkovic, S., Fetahovic, I., Mekic, E., Katipoglu, G., & Utku, S. (2024). Innovative Approach to Teaching Distributed Systems inEducation 4.0. *INTERNATIONAL JOURNAL OF ENGINEERING EDUCATION*, *40*(5).

[3] Muslić, E., Mekić, E., Novalić, F., Simović, V., & Katipoğlu, G. Educational–Industry Knowledge Graphs for Curriculum and Industry Alignment: A Case Study in Artificial Intelligence–Driven Digital Image Processing. *Available at SSRN 6619782*.

[4] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., … OpenAI. (2023). *GPT-4 technical report*. *arXiv*. https://arxiv.org/abs/2303.08774

[5] Barone IV, C. R., Serafin, R., Ali-Gombe, A., Richard III, G. G., & Case, A. (2022). *A reverse engineering education needs analysis survey*. *arXiv*. https://arxiv.org/abs/2212.07531

[6] Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

[7] Chikofsky, E. J., & Cross II, J. H. (1990). Reverse engineering and design recovery: A taxonomy. *IEEE Software, 7*(1), 13–17. https://doi.org/10.1109/52.43044

[8] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs. (2009). *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009L0024

[9] GitHub Docs. (2026). *Responsible use of GitHub Copilot features*. https://docs.github.com/en/copilot/responsible-use-of-github-copilot-features

[10] He, J., & Vechev, M. (2023). Large language models for code: Security hardening and adversarial testing. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. https://arxiv.org/abs/2302.05319

[11] He, J., Vero, M., Krasnopolska, G., & Vechev, M.

(2024). *Instruction tuning for secure code generation*. *arXiv*. https://arxiv.org/abs/2402.09497

[12] Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Final report of the Industrie 4.0 Working Group*. acatech. https://en.acatech.de/publication/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/

[13] Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering, 6*(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4

[14] Lyu, W., Wang, Y., Chung, T. R., Sun, Y., & Zhang, Y. (2024). Evaluating the effectiveness of LLMs in introductory computer science education: A semester-long field study. In *Proceedings of the ACM Conference on Learning @ Scale*. https://doi.org/10.1145/3657604.3662037

[15] National Institute of Standards and Technology (NIST). (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

[16] National Institute of Standards and Technology (NIST). (2026). *NICE Framework Resource Center*. https://www.nist.gov/itl/applied-cybersecurity/nice/nice-framework-resource-center

[17] Nie, A., Chandak, Y., Suzara, M., Malik, A., Woodrow, J., Peng, M., Sahami, M., Brunskill, E., & Piech, C. (2025). The GPT surprise: Offering large language model chat in a massive coding class reduced engagement but increased adopters' exam performances. In *Proceedings of the ACM Conference on Learning @ Scale*. https://arxiv.org/abs/2407.09975

[18] OWASP Foundation. (2025). *OWASP Top 10:2025*. https://owasp.org/www-project-top-ten/

[19] Mekic, E., Djokic, I., Zejnelagic, S. and Matovic A. (2016), "Constructive approach in teaching of VoIP in line with good laboratory and manufacturing practice. Computer Applications in Engineering Education, ID CAE-15-117.R1, 24: 277–287. doi: 10.1002/cae.21705

[20] Souppaya, M., Scarfone, K., & Dodson, D. (2022). *Secure software development framework (SSDF) version 1.1: Recommendations for mitigating the risk of software vulnerabilities* (NIST SP 800-218). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-218

[21] Votipka, D., Rabin, S. M., Micinski, K., Foster, J. S., & Mazurek, M. L. (2019). An observational investigation of reverse engineers' processes. In *28th USENIX Security Symposium (USENIX Security 19)* (pp. 1875–1892). https://www.usenix.org/conference/usenixsecurity19/presentation/votipka

[22] Mekić, Edis, Irfan Fetahović, Kristijan Kuk, Brankica Popović, and Petar Čisar. "Data-driven educational decision-making model for curriculum optimization." In *Ethics in Online AI-based Systems*, pp. 97-118. Academic Press, 2024.

[23] Stojanović, M., Matović, A., Matović, M., Мекić, E., Baždar, V., &Maamar, Z. (2025). "Enhancing Al algorithm in Digital Signal Processing Education through Extreme Programming: Constructive Approach in IoP context". *International Journal of Engineering Education*,Vol.41,No.6,pp.1-15,2025, https://www.ijee.ie/contents/c410625.html 1.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MRC COMBINER WITH THREE INPUTS IN THE PRESENCE OF FAST RICE FADING

ANA MATOVIĆ
Military Quality Control, Belgrade, ana.matovic4@gmail.com

MARIJA MATOVIĆ
Electrodistribution, Kraljevo, marija.matovic@es.rs

ELMEDIN BIBEROVIĆ
Telecom Serbia , N.Pazar, elmedin@telekom.rs

Abstract: *This paper analyzes a three-branch maximal-ratio-combining (MRC) receiver operating over independent and identically distributed fast, flat Rician fading channels. The model is stated directly in terms of the instantaneous branch signal-to-noise ratios (SNRs), which removes the ambiguity between Rician and Rayleigh fading in the original formulation. For ideal coherent MRC with perfect channel-state information and equal noise power in all branches, the output SNR is the sum of three independent Rician branch SNRs. This sum is a scaled noncentral chi-square random variable with six degrees of freedom and noncentrality parameter. Closed-form expressions are given for the output-SNR probability density function, cumulative distribution function, moment-generating and characteristic functions, and arbitrary raw moments. These results are then used to evaluate threshold-outage probability, ergodic spectral efficiency, coherent BPSK bit-error probability, noncoherent orthogonal BFSK error probability, and (M)-PSK symbol-error probability. Numerical results, supported by Monte Carlo validation, show the third-order diversity behavior of the three-branch receiver and quantify the reduction in deep-fade probability as the Rician (K)-factor increases at fixed average branch SNR.*

Keywords: *maximal-ratio combining; Rician fading; diversity; noncentral chi-square distribution; outage probability; bit-error probability; ergodic capacity.*

1. INTRODUCTION

Using the statistical characteristics of first-order systems—specifically, the system error probability, signal failure probability, average signal value, and channel capacity—we analyze the performance of a wireless telecommunication system. The failure probability is defined as the probability that the signal falls below a predefined failure threshold. This failure threshold is determined based on the required quality of service [12]. We also determine the channel capacity, i.e., the amount of information that can be transmitted from the transmitter to the receiver in a given time interval. In this study, we calculate the error probability for coherent systems, non-coherent systems, and systems using quadrature modulation [13].

Diversity techniques are commonly used to improve the performance of wireless signal transmission systems in the presence of various types of interference. In mobile telephony, spatial diversity is most often employed, using multiple antennas and combining received signals at the output of a combiner. In this study, we consider a wireless digital telecommunication system that employs maximal ratio combining (MRC) as a diversity technique to reduce the impact of fast Rician fading on system performance [4][5]. In an MRC receiver, the squared magnitude of the combined output signal equals the sum of the squared magnitudes of the signals at each input branch.

Specifically, we analyze an MRC combiner with three input branches under fast Rician fading. The Rician distribution models rapid fluctuations of the signal envelope in a line-of-sight environment [7-12]. A Rician random variable's square is the sum of the squares of two independent Gaussian random variables with equal variance, at least one having a non-zero mean.

In this case, the output of the three-branch MRC combiner is the sum of three squared Rayleigh-distributed random variables (equivalently, the sum of six squared Gaussian variables). As a result, the combined output signal is a random variable. We derive the probability density function (PDF) of the MRC output in equation (11). This PDF is used to compute the system's error probability (equation 24). Using the cumulative distribution function (equation 17) of the MRC output, we determine the probability of system failure (equation 23).

For completeness, we derive the moments of the MRC output signal: the first moment (mean), the second moment (average power, see equation 20), the third moment, and the general n-th moment. Finally, the channel capacity is computed as the expected value of Shannon's capacity expression (equation 22), representing

Corresponding author: A. Matović;
E-mail: ana.matovic4@gmail.com
DOI: 10.5937/OTEH26095M

the average information transmitted per unit time.

2. MRC COMBINER WITH THREE INPUTS IN THE PRESENCE OF FAST RICE FADING

2.1 Probability density of MRC combiner with three inputs in the presence of fast Rice fading

The probability densities of x_1, x_2 and x_3 are:

$$p_{x_1}(x_1) = \frac{2x_1}{\Omega} e^{-\frac{x_1^2+A^2}{\Omega}} I_0\left(\frac{2Ax_1}{\Omega}\right), \ x_1 \geq 0 \tag{1}$$

$$p_{x_2}(x_2) = \frac{2x_2}{\Omega} e^{-\frac{x_2^2+A^2}{\Omega}} I_0\left(\frac{2Ax_2}{\Omega}\right), \ x_2 \geq 0 \tag{2}$$

$$p_{x_3}(x_3) = \frac{2x_3}{\Omega} e^{-\frac{x_3^2+A^2}{\Omega}} I_0\left(\frac{2Ax_3}{\Omega}\right), \ x_3 \geq 0 \tag{3}$$

The signal x is equal to:

$$x = x_1^2 + x_2^2 + x_3^2 \tag{4}$$

The signal of the variable z is:

$$z = \frac{x}{\sigma^2}, \ x = z \cdot \sigma^2 \tag{5}$$

The probability density of z is:

$$p_z(z) = \frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{1}{4}} e^{-\frac{\lambda+z}{2}} I_{1/2}\left((z\lambda)^{1/2}\right) \tag{6}$$

where is: $\lambda = \frac{3A^2}{\sigma^2}, \ \sigma^2 = \frac{\Omega}{3}, \ \lambda = \frac{A^2}{\Omega}$ (7)

The probability density of x is:

$$p_x(x) = \left|\frac{dz}{dy}\right| p_z\left(\frac{x}{\sigma^2}\right) \tag{8}$$

where is:

$$\frac{dz}{dy} = \frac{1}{\sigma^2} \tag{9}$$

By substituting (6) into (9) in (8) we get:

$$p_x(x) = \frac{1}{2\sigma^2\sqrt[4]{3A^2}}(x)^{\frac{1}{4}} e^{-\frac{3A^2+x}{3\sigma^2}} I_{1/2}\left(\frac{A\sqrt{3x}}{\sigma^2}\right) \tag{10}$$

The probability density of x can be written in the form:

$$p_x(x) = \frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2+x}{3\sigma}} I_{1/2}\left(\frac{A\sqrt{3}\,x^{1/2}}{\sigma^2}\right) =$$
$$\frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2+x}{2\sigma^2}} \sum_{i=0}^{\infty}\left(\frac{A\sqrt{3}}{2\sigma^2}\right)^{2i+1/2} \frac{1}{i!\Gamma(i+3/2)} x^{\frac{1}{4}} =$$
$$a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\ x^{i+\frac{1}{4}} \tag{11}$$

where is:

$$a = \frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2}{2\sigma^2}}$$

$$\alpha = \frac{1}{2\sigma^2}$$

$$f_3(A,\sigma,i) = \left(\frac{A\sqrt{3}}{2\sigma^2}\right)^{2i+1/2} \frac{1}{i!\Gamma(i+3/2)} \tag{12}$$

The probability density of x can be written in the form:

$$p(a) = \frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2+x}{3\sigma}} I_{1/2}\left(\frac{A\sqrt{3}\,x^{1/2}}{\sigma^2}\right) =$$
$$\frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2+x}{2\sigma^2}} \sum_{i=0}^{\infty}\left(\frac{A\sqrt{3}}{2\sigma^2}\right)^{2i+1/2} \frac{1}{i!\Gamma(i+3/2)} x^{\frac{1}{4}} =$$
$$a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\ x^{i+\frac{1}{4}} \tag{13}$$

where is:

$$a = \frac{1}{2\sigma^2\sqrt[4]{3A^2}} x^{\frac{1}{4}} e^{-\frac{3A^2}{2\sigma^2}}, \alpha = \frac{1}{2\sigma^2} \tag{14}$$

$$f_3(A,\sigma,i) = \left(\frac{A\sqrt{3}}{2\sigma^2}\right)^{2i+1/2} \frac{1}{i!\Gamma(i+3/2)} \tag{15}$$

The characteristic function of x is:

$$\mu_x(s) = \overline{e^{xs}} = \int_0^{\infty} e^{xs} f_x(x)\,dx =$$
$$\int_0^{\infty} e^{xs}\, a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\ x^{i+\frac{1}{4}} dx =$$
$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \int_0^{\infty} \sum_{r=0}^{\infty} \frac{s^k x^k}{k!}\, x^{\frac{1}{4}} e^{-\alpha x}\, x^{i+\frac{1}{4}} dx =$$
$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \sum_{r=0}^{\infty} \frac{s^k}{k!} \int_0^{\infty} x^{k+i+\frac{1}{2}} e^{-\alpha x} dx =$$
$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \sum_{r=0}^{\infty} \frac{s^k}{k!} \left(\frac{1}{\alpha}\right)^{k+i+\frac{1}{2}} \Gamma(k+i+3/2) \tag{16}$$

2.2 Cumulative probability of MRC combiner with three inputs in the presence of fast Rice fading

The cumulative probability of a random variable x is:

$$F_x(x) = \int_0^x dt\, f_x(t) =$$
$$\frac{1}{2\sigma^2\sqrt[4]{3A^2}} e^{-\frac{3A^2}{2\sigma^2}} \sum_{i=0}^{\infty}\left(\frac{A\sqrt{3x}}{\sigma^2}\right)^{2i+1/2} \cdot$$
$$\frac{1}{i!\Gamma(i+3/2)} (2\sigma^2)^{i+3/2} \Gamma\left(i+3/2, \frac{x}{\sigma^2}\right) \tag{17}$$

2.3 Moments of MRC combiner with three inputs in the presence of fast Rice fading

The n-ti order moment of x is:

$$m_n = \overline{x^n} = \int_0^{\infty} x^n p_x(x)\,dx =$$
$$\int_0^{\infty} x^n\, a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\ x^{i+\frac{1}{4}} dx =$$
$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \int_0^{\infty} x^{n+i+\frac{1}{2}} e^{-\alpha x} dx =$$
$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \left(\frac{1}{\alpha}\right)^{n+i+\frac{1}{2}} \Gamma(n+i+3/2) \tag{18}$$

For n=1, we get the first moment, that is, the mean value of *x*:

$$m_1 = \bar{x} = \int_0^{\infty} x\, p_x(x)\, dx =$$

$$a \sum_{i=0}^{\infty} f_0(A,\sigma,i) \left(\frac{1}{\alpha}\right)^{i+\frac{5}{2}} \Gamma\left(i + 5/2\right) \tag{19}$$

For n=2, we get the second moment, or the mean square value of *x*:

$$m_2 = \overline{x^2} = \int_0^{\infty} x^2 p_x(x)\, dx =$$

$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \left(\frac{1}{\alpha}\right)^{i+\frac{7}{2}} \Gamma\left(i + 7/2\right) \tag{20}$$

For n=3, we get the third moment, or the average cubic value of *x*:

$$m_3 = \overline{x^3} = \int_0^{\infty} x^3 p_x(x)\, dx =$$

$$a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \left(\frac{1}{\alpha}\right)^{i+\frac{7}{2}} \Gamma\left(i + 7/2\right) -$$

$$\left(a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \left(\frac{1}{\alpha}\right)^{i+\frac{5}{2}} \Gamma\left(i + 5/2\right) \right)^2 \tag{21}$$

2.4 Capacity of the channel MRC combiner with three inputs in the presence of fast Rice fading

Channel capacity:

$$c = \int_0^{\infty} \left(ln(1+x^2)\right) p_x(x)\, dx = \int_0^{\infty} \left(ln(1+x^2)\right) a\, x^{\frac{1}{4}} e^{-\alpha x}$$

$$\sum_{i=0}^{\infty} f_3(A,\sigma,i)\, x^{i+\frac{1}{2}} dx =$$

$$\sum_{i=0}^{\infty} f_3(A,\sigma,i) \int_0^{\infty} x^{i+\frac{1}{2}} e^{-\alpha x} \left(ln(1+x^2)\right) dx \tag{22}$$

2.5 Probability of signal failure of MRC combiner with three inputs in the presence of fast Rice fading

The probability of signal failure is:

$$p_0 = \int_0^{x_0} p_x(x)\, dx = F_x(x_0) \tag{23}$$

2.6 Error probability of systems of MRC combiner with three inputs in the presence of fast Rice fading

Error probability for coherent systems:

$$p_{e_1} = \int_0^{\infty} a_1 \left(erfc(bx^2)\right) p_x(x)\, dx = \int_0^{\infty} a_1 \left(erfc(bx^2)\right)$$

$$a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\, x^{i+\frac{1}{2}} dx =$$

$$a_1 a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \int_0^{\infty} \left(erfc(bx^2)\right) x^{i+\frac{i}{2}} e^{-\alpha x} dx \tag{24}$$

Error probability for incoherent systems:

$$p_{e_2} = \int_0^{\infty} a_1 e^{bx^2} p_x(x)\, dx = \int_0^{\infty} a_1 e^{bx^2}$$

$$a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\, x^{i+\frac{1}{4}} dx =$$

$$a_1 a \sum_{i=0}^{\infty} f_3(A,\sigma,i) \int_0^{\infty} x^{i+\frac{i}{2}} e^{-\alpha x} e^{bx^2} dx \tag{25}$$

Error probability for quadrature modulated signals is:

$$p_{e_2} = \int_0^{\infty} dx \left(a_1 erfc(b_1 x^2) + a_2 \left(erfc(b_2 x^2)\right) p_x(x) =\right.$$

$$\int_0^{\infty} dx \left(a_1 erfc(b_1 x^2) + a_2 \left(erfc(b_2 x^2)\right)\right)$$

$$a\, x^{\frac{1}{4}} e^{-\alpha x} \sum_{i=0}^{\infty} f_3(A,\sigma,i)\, x^{i+\frac{1}{4}} \tag{26}$$

3. NUMERICAL RESULTS

Figure 1 shows the dependence of the probability density at the output of an MRC combiner with three inputs *p(x)* on the signal envelope *x*. For smaller values of the signal envelope *x*, the probability density *p(x)* increases, while for larger values of the signal envelope *x*, the probability density *p(x)* decreases. For a signal envelope *x* value from 1 to 7, the probability density *p(x)* increases from 0.001 to 0.09, while for a signal envelope *x* value from 7 to 20, the probability density *p(x)* decreases from 0.09 to 0.01.

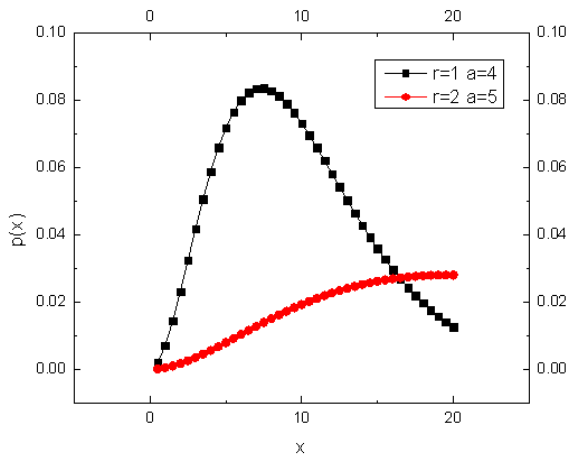


Figure 1. Dependence of the probability density at the output of the MRC combiner with three inputs on the parameter *x*

Figure 2 shows the dependence of the probability density at the output of an MRC combiner with three inputs on the parameter. For smaller values of the parameter, the probability density increases, while for larger values of the parameter, the probability density decreases. For parameter values from 0 to 5, the probability density rises from 0.048 to 0.078, while for parameter values from 5 to 10, the probability density decreases from 0.078 to 0.052.

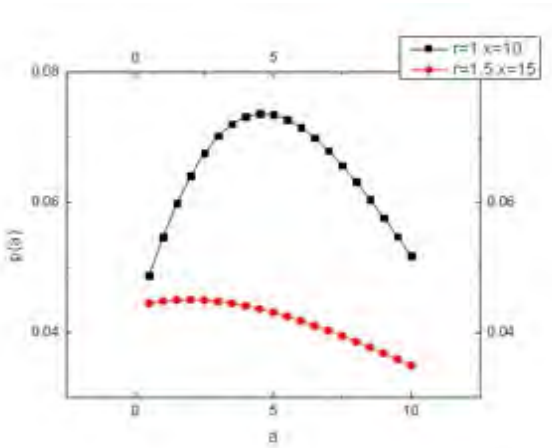


Figure 2. Dependence of the probability density at the output of an MRC combiner with three inputs on the parameter

Figure 3 shows the dependence of the mean square value of the signal at the output of the MRC combiner with three inputs, m_2, on the parameter a. For higher values of the parameter a, the mean square value of the signal m_2 increases. For a parameter a ranging from 0 to 10, the mean square value of the signal m_2 rises from 50 to 250.

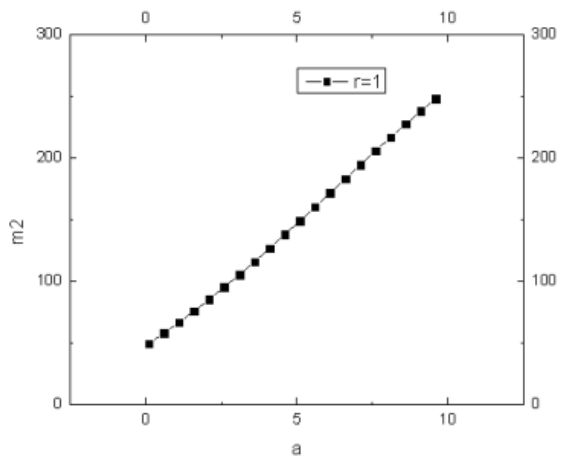


Figure 3. Dependence of the mean square value of the signal at the output of the MRC combiner with three inputs m_2 on the parameter a

4. CONCLUSION

This paper investigates an MRC combiner with three inputs in the presence of fast Rician fading. The probability density function and cumulative distribution function of the signal at the output of the three-input MRC combiner, as well as its first, second, and third moments, were derived under fast Rician fading conditions. Based on these results, the first-order statistical performance characteristics of the system were determined, including the system error probability, signal failure probability, average signal value, and channel capacity. Using the derived analytical expressions, the performance of the wireless communication system was analyzed, demonstrating the applicability of the MRC diversity technique for reducing the adverse effects of fast fading on system performance.

References

[1] M. K. Simon, M. S. Alouni, Digital Communications over Fading Channels: A Unified approach to performance analysis, John Wiley, New York, 2000.

[2] J. Proakis, Digital Communications, 3rd Ed., McGraw-Hill, 1999.

[3] Simon MK. Comments on "Infinite-series representations associated with the bivariate Rician distribution and their applications". IEEE Trans Commun series 2006;54 (8):1511-2.

[4] A. A. Abu-Dayya and N. C. Beaulieu, "Switched diversity on microcellular Ricean channels", IEEE Trans. Veh. Technol., vol.43 pp. 970-976, (1994).

[5] C. Tellambura, A. Annamalai and V. K. Bhargava, “Unified analysis of switched diversity systems in independent and correlated fading channels”, IEEE Trans. Commun., vol. 49, pp. 1955-1965, (2001).

[6] Časlav Stefanović, Marija Matović, Danijel Đošić, Srđan Maričić, Ana Matović, "Level crossing rate of the random variables", INFOTEH-JAHORINA, Vol. 10, Ref. B-I-4, ISBN 978-99938-624-6-8, Mart 20-

23, 2013.

[7] L.Wang and C.Lea, " co- channel interference analysis of shadowed Rician channels " IEEE Commun. Lett., vol. 2,no.3, pp.67-69, Mart 1999.

[8] Zogas DA, Karagiannidis GK. Infinite-series representations associated with the bivariate Rician distribution and their applications. IEEE Trans Commun series 2005;53 (11):1790-4.

[9] Mendes JR, Yacoub MD. A general bivariate Rician model and its statistics. IEEE Trans Commun series 2007;56(2):404-15.

[10] Bandjur, D.V., Stefanovic, M.C. and Bandjur, M.V. (2008).Performance analysis of SSC diversity receivers over correlated Ricean fading channels in the presence of co-channel interference. Electronic letters, vol.44,no.9, pp.587-588.

[11] Marko Smilić, Dusan Stefanović, Branimir Jaksić, Sinisa Minić, Edis Mekić, Ana Matović "Output Signal Moments of SC Macrodiversity with Two Microdiversity EGC Receivers", Microdiversity EGCReceivers", http://iris.elf.stuba.sk/JEEEC/data/pdf/6_115-09.pdf

[12] F.Vatalaro and G.E.Corazza, "Probability of error and outage in a Rician lognormal Channel for terrestrial and satellite personal communications," IEEE Trans. Commun., vol. 44, no. 8, pp. 921-924, Aug. 1996.

[13] M. S. Alouini, “A New Look at the Performance Analysis of Diversity Systems over Fading Channels: The MGF-Based Approach”, Trondheim, Norway, September 7, (1999).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

QUANTITATIVE ASSESSMENT OF ZERO TRUST ARCHITECTURE EFFECTIVENESS AGAINST ADVANCED CYBER ATTACKS IN A SMALL-SCALE EMULATED GOVERNMENT AND MILITARY INFORMATION SYSTEMS

BOBAN STOJANOVIĆ
Serbian Armed Forces, Belgrade, boban.stojanovic@vs.rs

MLADEN VEINOVIĆ
Singidunum University, Belgrade, mveinovic@singidunum.ac.rs

MARKO PREMOVIĆ
Serbian Armed Forces, Belgrade, marko.premovic@vs.rs

IGOR PETROVIĆ
Ministry of Defense, Belgrade, igor.petrovic@mod.gov.rs

Abstract: *In government and military networks, a successful intrusion does not necessarily depend on breaking through the network perimeter. APT actors may obtain valid credentials and use them to gain additional privileges and move between systems. We compared two security configurations in a small testbed containing 28 hosts. We used the same six MITRE ATT&CK techniques in both environments. For each run, we measured detection and response time, attacker dwell time, lateral movement, compromised hosts, and data exfiltration. We derived Attack Containment Efficiency (ACE) from the number of compromised hosts and used it as an additional containment measure. Detection in the traditional environment averaged 38.4 min, whereas in the Zero Trust environment it averaged 6.9 min. Response time fell from 71.8 to 14.2 min. Following the initial compromise, we observed an average of 12.4 successful lateral transitions per run in the traditional environment and 1.7 in the Zero Trust environment. The mean number of compromised hosts was 23 and 3, respectively, while successful data exfiltration decreased from 78% to 11%. Paired analysis showed significant differences between the configurations for all evaluated variables ($p < 0.001$). These results reflect the combined effect of the controls implemented in the Zero Trust-oriented configuration and should not be interpreted as the effect of any single security control.*

Keywords: *Zero Trust, MITRE ATT&CK, Military Systems, Attack Containment, Lateral Movement*

1. INTRODUCTION

Malware is not the only way into a military network. An attacker who obtains valid user credentials may already have the access needed to establish a foothold. From there, the account can be used to seek additional privileges, retain access, or reach other systems [1]. Identity-based attacks have also been observed in public-sector networks [2]. In a military setting, this kind of access can have wider consequences when operational or administrative systems supporting mission activities are affected [3].

Traditional network security largely focuses on protecting the boundary between trusted and untrusted networks [4]. This approach becomes less useful once an attacker has crossed that boundary. An attacker with internal access may attempt to reach other hosts and services from inside the network. NIST SP 800-207 therefore treats network location as insufficient grounds for trust and considers threats that may be present both inside and outside the traditional perimeter [5].

Zero Trust does not assume that a user or device can be trusted simply because it is already inside the network. Access decisions are instead based on identity and other available security information [4], [6]. An access request must therefore be evaluated without relying on network location as proof of trust. This approach is also reflected in NSA guidance, which presents Zero Trust as a means of maintaining authorized users' access to mission data while limiting an adversary's access when part of the network has been compromised [4].

This study was motivated by the lack of quantitative evidence on Zero Trust effectiveness in military networks. We built two matched test environments with the same host and application inventory but different security configurations, ran the same attacks against both, and compared the results.

1.1 Related Work

Kindervag [6] proposed Zero Trust as a different way of thinking about network security. Instead of assuming that systems behind the perimeter are trustworthy, access must be considered separately for each resource. NIST SP 800-207 [5] later described this idea in greater technical detail,

Corresponding author: Boban Stojanović;
E-mail: boban.stojanovic@vs.rs
DOI: 10.5937/OTEH26096S

specifying the components responsible for making and enforcing access decisions. Experimental studies have since examined micro-segmentation, including work based on network emulation and attack graphs.

Controlled attack simulations provide one way to examine how security controls behave during an intrusion. MITRE ATT&CK [7] provides the terminology for describing the adversary techniques used in such tests. Caldera can execute the same ATT&CK techniques in an emulated environment, allowing detection and response to be observed under controlled conditions [7]. Hutchins et al. [8] earlier structured an intrusion using the intrusion kill chain model, dividing an adversary campaign into stages.

Attack graphs have been used to examine how far an attacker can move through a segmented network [9]. Other researchers have tested micro-segmentation by reproducing attacks in controlled environments [10]. Most of this work, however, deals with enterprise networks. Military-oriented test environments are less frequently examined, and containment is rarely expressed through a metric such as ACE.

This study addresses that gap by evaluating Zero Trust Architecture against six MITRE ATT&CK techniques in a small-scale emulated military environment, using paired runs and a controlled experimental design. The selected post-compromise behaviors are also consistent with APT techniques summarized in the literature [11].

2. RESEARCH METHODOLOGY

2.1 Experimental Design

The experimental setup consisted of two matched environments running on a VMware vSphere cluster. The hardware and software inventory was kept the same in both cases: 25 Windows 10 endpoints, three Windows Server 2022 systems providing file, application, and database services, an Active Directory domain, and Splunk for centralized logging. This gave 28 hosts in each environment. The environments differed in the security configuration applied to them.

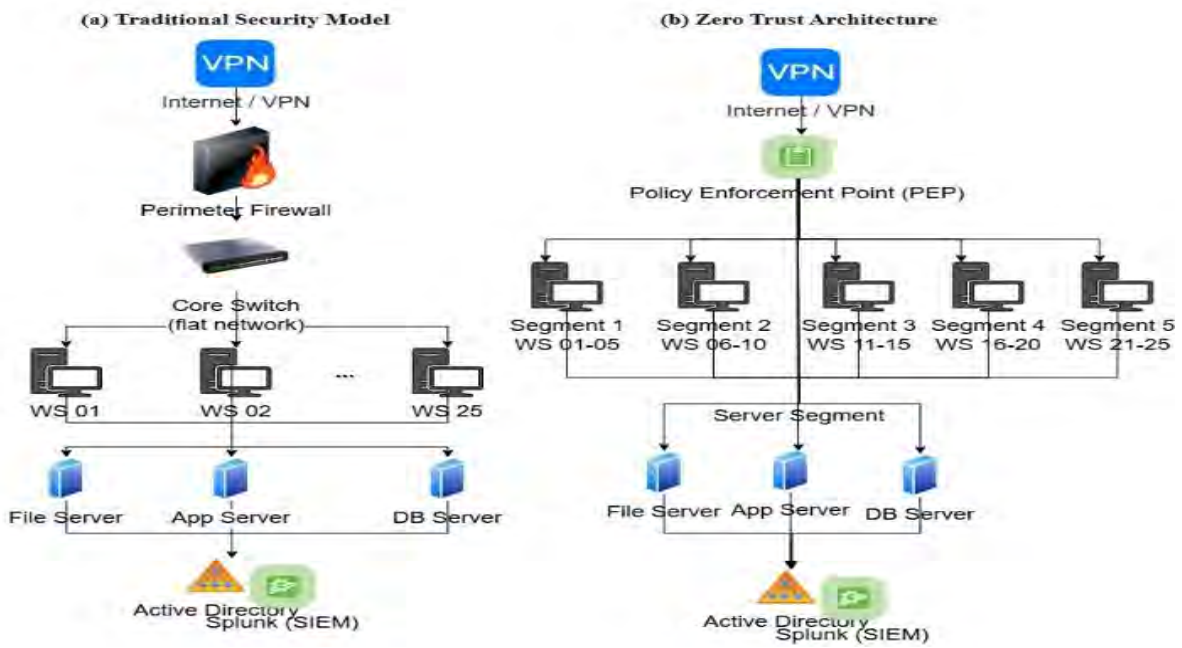


Figure 1. Topology of the two emulated environments: (a) Traditional Security Model with perimeter-only protection and implicit trust between hosts; (b) Zero Trust Architecture with five micro-segments, policy enforcement points (PEP), identity-aware access control, and least-privilege enforcement.

Environment A – Traditional Security Model: Flat internal network topology; perimeter firewall protection; VPN-based remote access; static authentication (password + MFA for external access); domain-based trust model.

Environment B – Zero Trust Architecture: We split the network into five segments and applied ABAC to control access between them. We included device posture in the access decision. We required re-authentication every 60 minutes and monitored user activity and active sessions between checks. Policy and user privileges limited access to resources [12].

2.2 Attack Simulation Scenarios

Six attack cases were executed in accordance with the MITRE ATT&CK Enterprise Framework [7]:

Table 1. MITRE ATT&CK Attack Cases

Attack Phase	Technique	Zero Trust Mitigation
Initial Access	T1078 - Valid Accounts	Periodic re-authentication + behavioral monitoring
Privilege Escalation	T1003 - Credential Dumping	Least-privilege access
Lateral Movement	T1550.002 - Pass-the-Hash	Network segmentation
Execution	T1059.001 - PowerShell	Application control
Lateral Movement	T1210 - Remote Service Exploit	Micro-segmentation
Propagation	T1570 - Lateral Tool Transfer	Traffic monitoring

An automated adversary emulation platform (MITRE Caldera) executed the attacks using predefined ability scripts mapped to the six ATT&CK techniques in Table 1. Each run executed a single technique against a freshly reverted snapshot. We reverted the environment to the baseline snapshot before each run and verified that Active Directory, the Splunk index, and the endpoint state were clean.

Both environments used the same Splunk-based SOC pipeline, with no changes to the correlation rules or alert thresholds. In Environment A, the analyst followed the standard response playbook and carried out the required actions manually. The same analyst handled the alerts in Environment B, but session termination and segment isolation could also be enforced automatically. Since the analyst knew which environment was being tested, the experiment was not blinded.

The first correlated alert recorded in Splunk marked the detection point for a run. We recorded containment after either isolating the compromised host or deactivating the account used in the attack. For LMS, we recorded each successful lateral transition during a run. DESR represented the share of the predefined canary files that successfully reached the external exfiltration listener.

We repeated each technique 20 times, for 120 runs per environment. We matched runs from the two environments by technique and initial access vector. Thus, both members of a pair started from the same access vector. This pairing allowed the two security configurations to be compared under equivalent starting conditions [10].

The security setup was not changed one control at a time. In Environment B, several controls were active simultaneously: network segmentation, ABAC, device-posture checks, application control, behavioral monitoring, periodic re-authentication, and policy enforcement. Environment A did not include these changes and retained the flat-network configuration. Consequently, the experiment compares the two implemented setups. It does not show which individual control drove a particular improvement. Answering that question would require additional tests in which the controls were introduced or removed separately.

2.3 Assessment Criteria

The following metrics were measured:

- **MTTD** – time from the initial compromise event to the first SOC alert.
- **MTTR** – time from the first SOC alert to the containment action.
- **ADT** – time from the initial compromise event to the eradication of the attacker's presence. ADT includes both the undetected and response periods.
- **LMS** – number of successful lateral transitions per run.
- **CH** – total number of systems compromised per run.
- **DESR** – percentage of the predefined canary file set successfully exfiltrated per run.

Note on MTTD and ADT. MTTD and ADT are not the same interval. MTTD ends at the first alert; ADT ends at eradication. In the traditional environment, the attacker remained active for 252 min because detection was late (38.4 min) and containment was slow (71.8 min); in the Zero Trust environment, ADT was 29 min because detection was fast (6.9 min) and containment was fast (14.2 min). The relationship ADT ≈ MTTD + MTTR + residual attacker activity is consistent with the recorded event sequence.

2.4 Attack Containment Efficiency (ACE)

ACE is a normalization of the compromised-host count to the interval [0, 1]:

$$ACE = 1 - \frac{CH}{TH} \tag{1}$$

where CH = number of compromised hosts and TH = total number of hosts (28). ACE values approaching 1 indicate stronger containment. ACE is a transformation of CH and TH already reported in Table 2.

2.5 Statistical Design

For every run in one environment, we selected the corresponding run in the other environment with the same ATT&CK technique and initial access vector. This produced 120 pairs for the analysis. We compared the values within these pairs using a paired t-test. We also checked the differences between paired observations with the Shapiro-Wilk test, and $p > 0.05$ for every variable. We obtained effect sizes directly from the paired-test statistic using Cohen's

$$d_z = t/\sqrt{n} \tag{2}$$

where $n = 120$ is the number of pairs. We interpreted the resulting $d_z$values using the benchmarks proposed by Sawilowsky [13].

3. EXPERIMENTAL RESULTS

Table 2. Security Performance Comparison (mean over 120 paired runs per environment)

Metric	Traditional	Zero Trust	Improvement
MTTD (min)	38.4	6.9	82.0% reduction
MTTR (min)	71.8	14.2	80.2% reduction
ADT (min)	252	29	88.5% reduction
Lateral Movements (count)	12.4	1.7	86.3% reduction
Compromised Hosts (count)	23	3	87.0% reduction
Data Exfiltration (%)	78	11	85.9% reduction

Changes were evident across all six measures. Attacker dwell time showed the largest reduction. The number of hosts reached during an attack also dropped sharply, and only a small number of lateral transitions were completed in the Zero Trust environment. The same pattern appeared in the exfiltration tests, where considerably less of the predefined test data left the environment. Some variation occurred between individual runs, as expected, but it did not alter the general pattern.

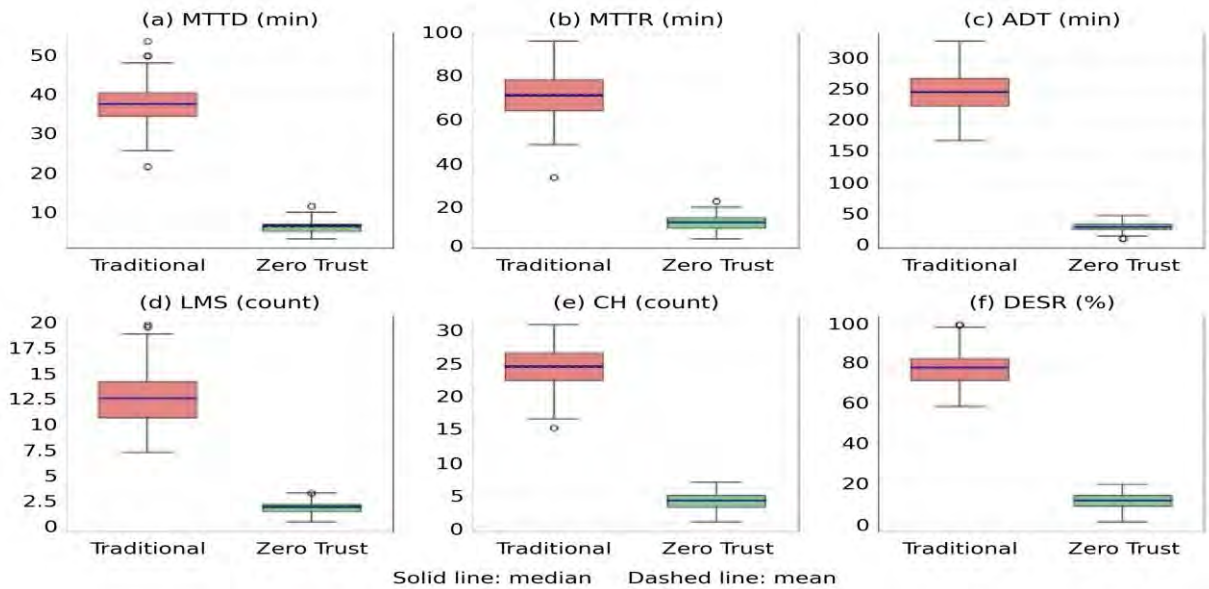


Figure 2. Comparison of the two environments over 120 paired runs. The plots present (a) MTTD, (b) MTTR, (c) ADT, (d) lateral movements, (e) compromised hosts, and (f) data exfiltration rate. The line inside each box marks the median, the box covers the middle 50% of observations, and the whiskers extend to 1.5 × IQR.

Attack detection took an average of 38.4 minutes in the traditional environment. With the controls used in Environment B, the corresponding time was 6.9 minutes. Response time followed the same pattern, decreasing from 71.8 to 14.2 minutes. This represents reductions of 82.0% and 80.2%, respectively. During the log review, we also found that containment generally started earlier in Environment B. Policy-violation alerts were one of the sources available to the analyst when deciding whether to initiate containment.

Dwell time showed the largest change. In the traditional setup, an attacker remained active for 252 min on average; in the Zero Trust setup, this value was 29 min (88.5% lower). A run in Environment A also produced 12.4 successful lateral transitions on average, compared with 1.7 in Environment B (86.3% lower). The mean number of affected hosts was 23 in the first environment and 3 in the second, corresponding to an 87.0% reduction. We observed the same direction of change for exfiltration: 78% of the predefined data set reached the external listener in Environment A, compared with 11% in Environment B. This is a 67-percentage-point difference, or an 85.9% relative reduction. Each comparison compared runs that shared the same ATT&CK technique and initial access vector.

4. STATISTICAL ANALYSIS

4.1 Paired-Samples t-Test

Table 3. Statistical significance results (paired runs, n = 120 per environment)

Variable	Traditional (mean ± SD)	Zero Trust (mean ± SD)	Mean diff.	SD(d)	t (df = 119)	p	Cohen's d_z
MTTD (min)	38.4 ± 6.2	6.9 ± 1.4	31.5	6.4	53.9	< 0.001	4.92
MTTR (min)	71.8 ± 11.5	14.2 ± 3.1	57.6	11.9	53.0	< 0.001	4.84
ADT (min)	252 ± 34	29 ± 7	223	34.7	70.4	< 0.001	6.43
LMS (count)	12.4 ± 2.8	1.7 ± 0.6	10.7	2.9	40.4	< 0.001	3.69
CH (count)	23 ± 3.1	3 ± 1.2	20.0	3.3	66.4	< 0.001	6.06
DESR (%)	78 ± 9	11 ± 4	67.0	9.9	74.1	< 0.001	6.77

The paired tests showed statistically significant differences between the two configurations ($p < 0.001$ for all evaluated variables). Under the null hypothesis of no paired mean difference, the reported p-values quantify how incompatible the observed test statistics are with that hypothesis. Cohen's d_z values derived from the paired t-statistics range from approximately 3.7 to 6.8, indicating large paired effects in the experimental dataset.

4.2 Pearson Correlation Analysis

We examined the relationship between detection time (MTTD) and attacker dwell time (ADT) across all runs. When we pooled all 240 runs, the correlation was $r = 0.87$ ($p < 0.001$). This pooled value partly reflects the large difference between the two environments rather than a within-environment relationship. When we analyzed each environment separately, the correlations were lower: $r = 0.42$ for the traditional setup and $r = 0.38$ for the Zero Trust setup. We therefore use the correlation results only to describe the observed data, not to infer causality.

5. CONTAINMENT ANALYSIS

5.1 Attack Containment Efficiency Calculation

Using the experimental data with TH = 28 (25 endpoints + 3 servers):

Traditional architecture:

$$ACE_{TRAD} = 1 - \frac{23}{28} = 0.179 \quad (3)$$

Zero Trust architecture:

$$ACE_{ZT} = 1 - \frac{3}{28} = 0.893 \quad (4)$$

ACE increased from 0.179 in Environment A to 0.893 in Environment B, giving a difference of 71.4 percentage points. This value comes directly from the compromised-host results in Table 2. We therefore use ACE only to express those results on a normalized scale, rather than treating it as a separate measurement.

6. DISCUSSION

The experimental results show that the Zero Trust-oriented configuration provided faster detection and containment and more effectively restricted post-compromise propagation than the traditional configuration under the evaluated ATT&CK scenarios.

Several security controls were active simultaneously in the Zero Trust environment, including micro-segmentation, ABAC, device-posture checks, application control, behavioral monitoring, periodic re-authentication, and policy enforcement. The measured improvement therefore cannot be assigned to any one of these controls. Our results describe the difference between the two configurations as they were tested.

We had to revise several rules while setting up the testbed. Some early rules restricted administrative tasks still necessary to operate the environment. After testing, we adjusted those rules so that the required tasks remained possible. This experience also showed a practical limitation of strict access control: security restrictions must be introduced without disrupting the administration of systems that need to remain available.

The 60-minute re-authentication requirement was likewise only one part of the setup: because a valid credential could still be abused between checks, behavioral monitoring, least-privilege restrictions, and session-termination mechanisms were used alongside it to identify suspicious activity and end a session when necessary.

Concrete Benefits for Military Systems:

- **Reduced Blast Radius:** Limiting the spread of a compromise can reduce the number of mission-supporting systems exposed after an attacker gains access.
- **Faster Incident Response:** Environment B combined the analyst's response workflow with automated session termination and segment isolation. The response-time comparison therefore reflects both the security controls and the response processes used in the testbed.
- **Operational Resilience:** In this setup, trust did not depend only on where a user or device was located in the network. Segmentation limited access between network areas, while identity policies and least-privilege rules restricted access to individual resources. Monitoring provided an additional check during active sessions. Together, these measures also allowed us to isolate an affected segment without necessarily disrupting services in other segments. This follows the basic idea of defense in depth [14].

7. LIMITATIONS OF THE STUDY

Our experiment compares the two environments as they were configured, not the effect of one security control in isolation. Several mechanisms used in Environment B were not present in Environment A, so we cannot assign the measured differences to any one of them. The experiment was also limited to 28 hosts and six ATT&CK techniques. We did not examine larger distributed networks, insider attacks, supply-chain compromise, DDIL operating conditions, or the performance cost of the implemented controls.

8. CONCLUSION

The two environments produced different results. In the 28-host testbed, Environment B detected attacks sooner and required less time to contain them. Attackers also completed fewer lateral transitions, reached fewer hosts, and exfiltrated less of the predefined test data. These results apply to Environment B as it was tested, with its security controls operating together. They do not show how much any one control contributed to the observed differences.

ACE expresses the compromised-host result on a normalized scale. Its value was 0.179 for Environment A and 0.893 for Environment B, a difference of 71.4 percentage points. The paired tests found significant differences for all variables included in the comparison. ATT&CK kept adversary actions organized consistently across runs and linked those actions to the observed defensive results.

These results come from testing all controls together. This experiment cannot determine which control produced which part of the difference. That would require another set of tests, changing one control at a time while leaving the rest of the setup unchanged.

A larger test environment would be a useful next step, particularly one that includes systems at more than one location. The attack set could also be expanded beyond the six ATT&CK techniques used here to cover insider activity, supply-chain compromise, and DDIL conditions. This experiment did not measure resource and performance costs, which should be included in further testing. Finally, testing the Zero Trust controls one at a time would help determine how much each control contributes to the observed results.

References

[1] VERIZON: 2024 Data Breach Investigations Report, Verizon Enterprise Solutions, 2024.

[2] CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY (CISA): Zero Trust Maturity Model, Version 2.0, CISA, Washington, DC, 2023.

[3] U.S. DEPARTMENT OF DEFENSE: DoD Zero Trust Strategy, U.S. Department of Defense, Washington, DC, 2022.

[4] NATIONAL SECURITY AGENCY: Embracing a Zero Trust Security Model, Cybersecurity Information Sheet, NSA, 2021.

[5] ROSE, S., BORCHERT, O., MITCHELL, S., CONNELLY, S.: Zero Trust Architecture, NIST Special Publication 800-207, National Institute of Standards and Technology, Gaithersburg, MD, 2020, doi: 10.6028/NIST.SP.800-207.

[6] KINDERVAG, J.: Build Security Into Your Network's DNA: The Zero Trust Network Architecture, Forrester Research, Cambridge, MA, 2012.

[7] MITRE CORPORATION: MITRE ATT&CK Enterprise Matrix, 2026. [Online]. Available: https://attack.mitre.org.

[8] HUTCHINS, E. M., CLOPPERT, M. J., AMIN, R. M.: Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains, Leading Issues in Information Warfare & Security Research, Vol. 1, No. 1, 2011, pp. 80–106.

[9] GAIN, A., BARIK, M. S.: Augmenting Zero Trust Architecture with Temporal Attack Graph, 2025 IEEE 6th India Council International Subsections Conference (INDISCON), Rourkela, India, 2025, pp. 1–6.

[10] FRYDMAN, C., ZILBERMAN, A., KRIEF, R., SHOWGAN, A., MURILLO, A., MOTOYOSHI, S., SHABTAI, A., ELOVICI, Y., PUZIS, R.: COHORT: Collaborative Orchestration for Hardening via Offensive Replay on Emulated Topologies, arXiv preprint arXiv:2606.30479, 2026.

[11] ALSHAMRANI, A., MYNENI, S., CHOWDHARY, A., HUANG, D.: A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities, IEEE Communications Surveys & Tutorials, 21(2) (2019) 1851-1877.

[12] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY: Security and Privacy Controls for Information Systems and Organizations, NIST Special Publication 800-53 Rev. 5, National Institute of Standards and Technology, Gaithersburg, MD, 2020, doi: 10.6028/NIST.SP.800-53r5.

[13] SAWILOWSKY, S. S.: New Effect Size Rules of Thumb, Journal of Modern Applied Statistical Methods, Vol. 8, No. 2, 2009, pp. 597–599, doi: 10.22237/jmasm/1257035100.

[14] STALLINGS, W., BROWN, L.: Computer Security: Principles and Practice, Pearson, Boston, MA, 2023.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CHARACTERIZATION AND SUPPRESSION OF COMMON-MODE AND DIFFERENTIAL-MODE CONDUCTED EMI

PAVEL CRNOMARKOVIĆ
Military technical institute, Belgrade, crnomarkovic.pavel@gmail.com

SAŠA GUNDELJ
Military Technical Institute, Belgrade, gundelj@yahoo.com

MIHAILO KNEŽEVIĆ
Faculty of Technical Sciences, Čačak, mihailo.knezevic@ftn.kg.ac.rs

RADOSLAV SURLA
Military Technical Institute, Belgrade, ekorade@gmail.com

Abstract: *Conducted electromagnetic interference (EMI) is an important aspect of electromagnetic compatibility (EMC), especially in electronics system where noise can propagate through power and signal lines, and affect of operation of other circuits. Proper identification of the dominant type of noise is therefore essential for the of effective EMI suppression. This paper presents measurement based methodology for the analysis and characterization of conducted EMI, with particular emphasis on the separation of differential mode (DM) and common mode (CM)components. The proposed approach includes the configuration of measurement equipment, acquisition and analysis of measured data, and identification of dominant frequency dependent noise component. Based on measurement results, total conducted interference is decomposed into its DM and CM component. The obtained results are used to select appropriate suppression for each noise mode. The presented methodology provides a systematic approach for distinguishing EMI mode for identification of suppression, thereby improving the effectiveness of EMC filtering.*

Keywords: *Electromagnetic compatibility, Conducted EMI, common-mode noise, differential-mode noise, characterization, EMI suppression.*

1. INTRODUCTION

The rapid development of modern electronics has led to a significant increase in the use of power electronic convertors, switching power supply, high speed digital circuits, and digitally controlled subsystem. DC/DC convertors, AC/DC power supply, motor drives, etc., have become integral parts of a wide range of electronic and electrical equipment. Although these technologies provide high efficiency, compact dimension, and improved functionality and working stability, their switching operation represents a significant source of EMI.

The EMC performance of a modern electronic devices is determined not only by its intended electrical functionality, but also by its ability to limit the generation and propagation of unwanted electromagnetic energy. EMC requirements have become an essential part of the design, and verification process of electronic equipment. A device intended for commercial, industrial, automotive, aerospace, military, or other specialized applications must satisfy the applicable EMC requirements before it can be placed on the market or introduced into service.

Practically all country, around the world, require that electrical and electronic devices must meet strict requirements regarding EMC. [1]

Various international and industry specific standard define limits and measurement procedures for electromagnetic emissions and immunity. Examples include the CISPR series of standards, IEC 61000-4series, and military standard MIL-STD-461. The specific requirements depend on intended application and environment of the equipment. In particular, conducted emission measurements are used to determine the level of unwanted RF energy present on the power inputs and output interfaces of equipment under test (EUT). In the presented paper, conducted emission measurements are performed in accordance with the requirements of the MIL-STD-461 CE102 test method, which specifies limits and measurements procedures for conducted emissions on power input leads of equipment and subsystems. [2]

A limitation of conventional conducted emission measurement is that the measured spectrum represents the combined effect of different noise type. The measured interface may contain both DM and CM components. Although these two modes may appear as a single conducted emission spectrum at the measurement port, they are completely different in term of appearance and propagation (it will be clarified in next chapter) and

Corresponding author: P. Crnomarković;
E-mail: crnomarkovic.pavel@gmail.com

DOI: 10.5937/OTEH26097C

require different suppression technique.

In this paper will be presented expansion of the MIL-STD-461 CE102 method in order to obtain more precise data on the type and level of interference in order to more precisely determine the parameter of filter for suppression, and way haw to choose right filter.

2. TEORETICAL BACKGROUND

Conventional conducted measurement represents both DM and CM noise Figure 1. The yellow line represents path of the sum of the CM and DM components, while the green line represents path of difference of CM and DM component, depending which port is used for measuring from line impedance network stabilizer (LISN), URF1 or URF2.

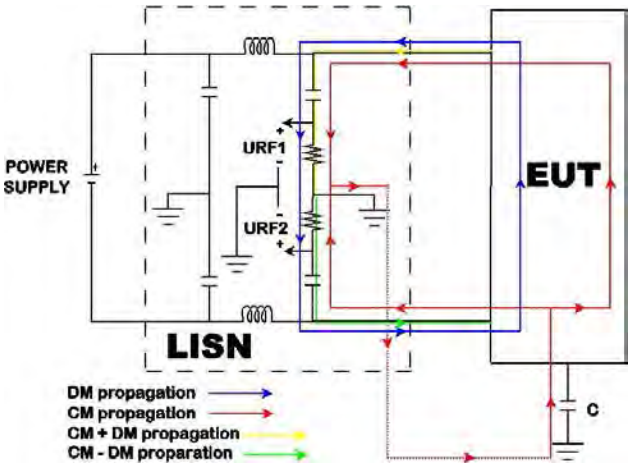


Figure 1. Conventional measurement

Identification and separation of DM and CM component represent an important step in the conducted EMI troubleshooting and process of choosing adequate filter. DM noise are associated with current flowing between power lines in opposite direction Figure 2, whereas CM noise are generally associated with currents flowing in same direction on power lines with respect to a chassis Figure 3.

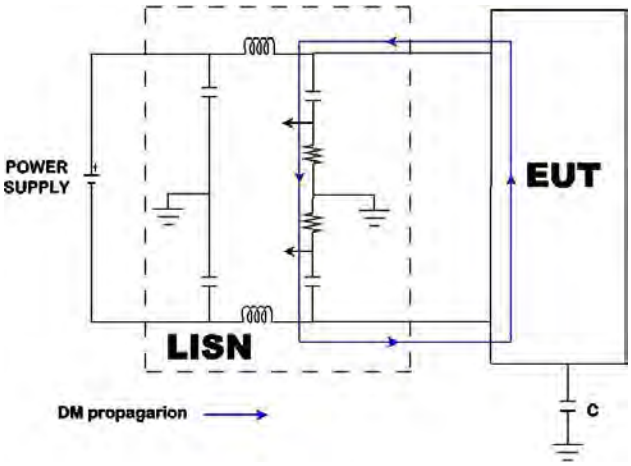


Figure 2. DM noise propagation

Mathematically speaking it is possible to distinguish U_{CM} and U_{DM} from URF1 and URF2 in following way:

$$U_{CM} = \frac{UREF1 + UREF2}{2} \quad (1)$$

$$U_{DM} = \frac{UREF1 - UREF2}{2} \quad (2)$$

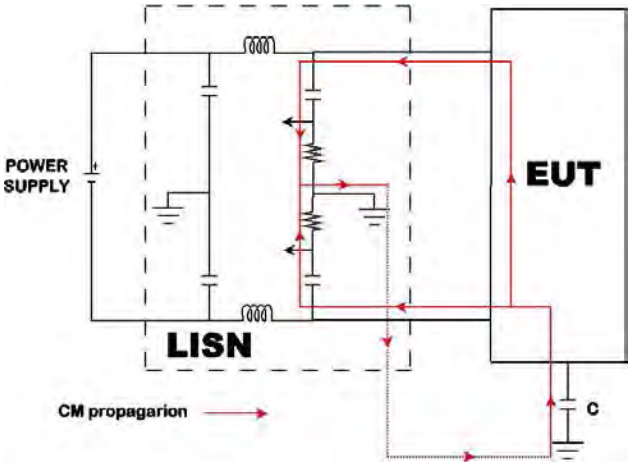


Figure 3. CM noise propagation

3. MEASUREMENT MODEL

The measurements were performed in EMC test chamber. General test setup was as proposed by MIL-STD-461 standard [2]. General test setup is shown on Figure 4. Same figure represents position of equipment distances of cable, size of ground plane, dimension of non-conductive standoff, separation of element relatively to the ground plane on the table and also to the ground of the EMC test chamber etc.

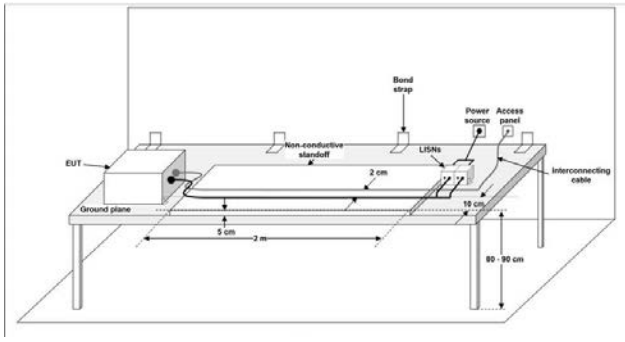


Figure 4. General test setup [2]

For MIL-STD-461, CE102 conducted emission measurement is used equipment setup as shown on Figure 5. The equipment are consists of 2 LISN unit, termination of 50ohm, which is used to terminate unused RF port from LISN unit, attenuation from 20dB, which is used to protect measuring equipment in case of appearance of very high level of conductive interference, EUT, measurement devices and data recording device. Lately last two devices are integrated in same unit. Also power supply should be galvanically isolated.

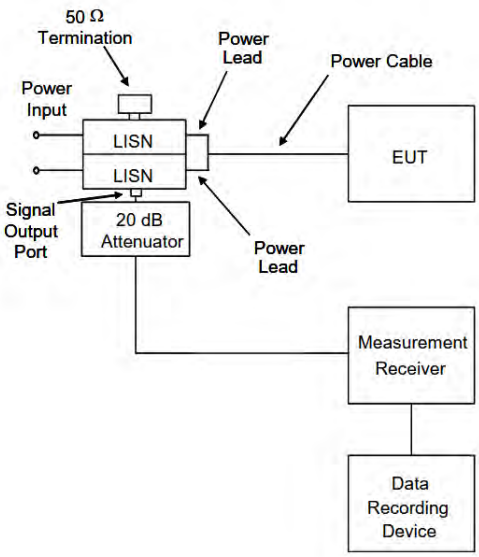


Figure 5. Measurement equipment setup [2]

Modification of measurement setup as proposed MIL-STD-461, CE102, consists of using MISN-MATE unit. This unit is feed from booth RF outputs, from each LISN unit, RF signal from LISN on positive power line is connected to the RF1 port of the LISN-MATE, and RF signal from LISN on negative power line is connected to the RF2 port of LISN-MATE unit.

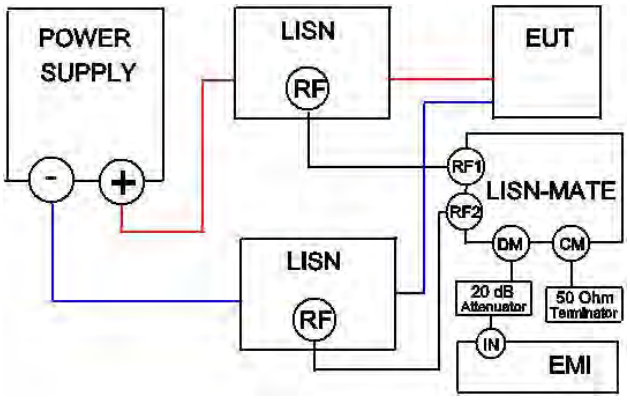


Figure 6. Expanded measurement equipment setup for DM component

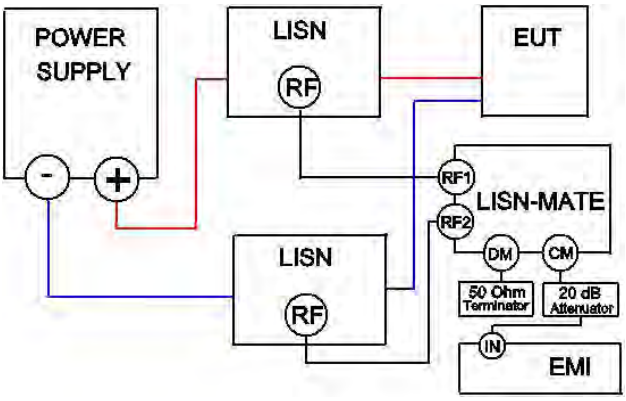


Figure 7. Expanded measurement equipment setup for CM component

LISN-MATE at its output provides DM signal output, which is calculated according to quotation (3), and CM signal output which is calculated according to quotation (4).

$$U_{CM} = URF1 + URF2 \tag{3}$$

$$U_{CM} = URF1 - URF2 \tag{4}$$

Comparing quotation (3) and (4) to quotation (1) and (2) respectively, is obvious that LISN-MATE provide twice the level on both outputs, so it is necessity to set attenuation of 3dB in measuring equipment, for compensation.

Proposed expanded measurement equipment setup is shown on Figure 6. and Figure 7., respectively.

Data presentation shall be as follows:

a. Continuously and automatically plot amplitude versus frequency profiles on X-Y axis outputs. Manually gathered data is not acceptable except for plot verification.
b. Display the applicable limit on each plot.
c. Provide a minimum frequency resolution of 1% or twice the measurement receiver bandwidth, whichever is less stringent, and a minimum amplitude resolution of 1 dB for each plot.
d. Provide plots for both the measurement system check and measurement portions of the procedure. [2]

The measurement system check is performed using the measurement system check chart for CE102, shown on Figure 8.

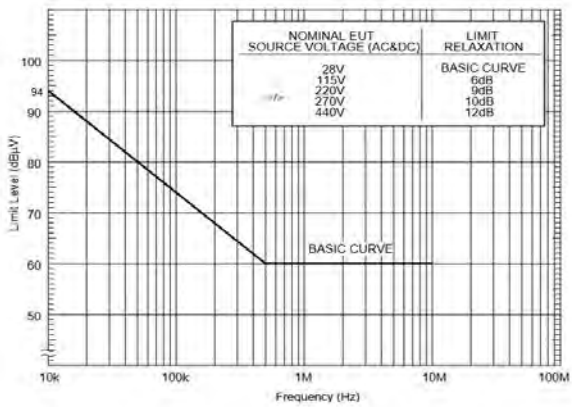


Figure 8. CE102 limit (EUT power leads, AC&DC) [2]

4. RESULTS

In this chapter, concrete measuring equipment, measurement results, a proposal for selecting a filter, and a verification measurement for confirmation will be presented.

4.1 Equipment

In conducted tests are used next equipment.

LISN is LI-3100 from COM-POWER, very powerful, up to 100Amps, can be used and for DC and for AC, and also very good quality, more detail is given at the reference [3]

LISN-MATE is TBLM2 from TEKBOX, with adequate parameters, and also very good quality, more details at the reference [4]

EMI from Agilent Technologies N9038A, nowadays is Keysight Technologies N9038A, it's a standards compliant EMI receiver and diagnostic signal analyzer built on an upgradeable platform. In the lab and on the bench, it provides the accuracy, repeatability, and reliability you need to test with confidence. [5]

Terminator 50 Ohm from HP, nowadays Keysight 909A coaxial terminator. Terminate a 50 Ω coaxial system in its characteristic impedance with this precision, low-reflection load. [6]

Attenuator 20dB from HP 8491B, nowadays Keysight Technologies, provides precision attenuation, flat frequency response, and low SWR over a broad frequency range. [7]

RF cables from HP very good quality, is good to mention, that cable for connection between LISNs and LISM-MATE unit should be from same manufacturer and same length, because difference in attenuation in this cable can affect this measurement and provides inaccurate data in difference in levels between CM and DM interference.

Power cables should be calculated according to maximum load of the EUT, and also cables length, between LISN and EUT must be 2 m long, as shown on Figure 4.

Power supply in our case was DC, and galvanically isolated.

EUT is multi rail switching power supply.

4.2 Equipment setup

For measurement is used setup as shown on Figure 6, for measurement of DM interference, and as shown on Figure 7, for CM interference.

EUT is set to 100% of load. Reason is that the switching power supply, what is EUT in this case, would generate the most interference at maximum load.

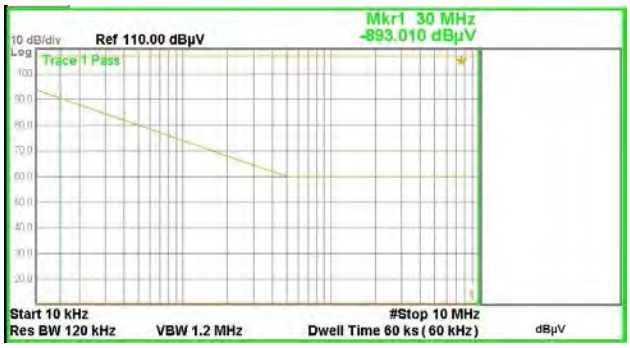


Figure 9. EMI state for CE102

Since we use EMI to collect and record data, the possibilities it offers are used, a state is created, (called state CE102) according to the requirements of data presentation and configuration of the equipment (Figure 6 and Figure 7). For the threshold curve according to the CE102 standard, Figure 8, basic curve is used, because power supply voltage was 24Vdc. Also, the attenuation of the attenuator from 20dB was entered, and an attenuation of 3dB was added due to LISN-MATE unit multiplication of output signal. This state is shown on Figure 9, before measurement.

4.3 Measurement

After calibration of equipment, all equipment were connected, as shown on figure 6, power supply is off, state CE102 is called and ambient noise is measured Figure 10.

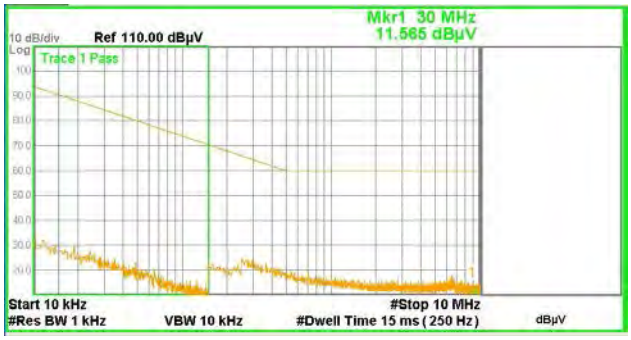


Figure 10. Ambient interference

4.3.1 DM component measurement

Difference to previous setup is that power supply was turn on, load is set to maximum load, everything else was the same, state CE102 was recalled and DM interference level was measured Figure 11.

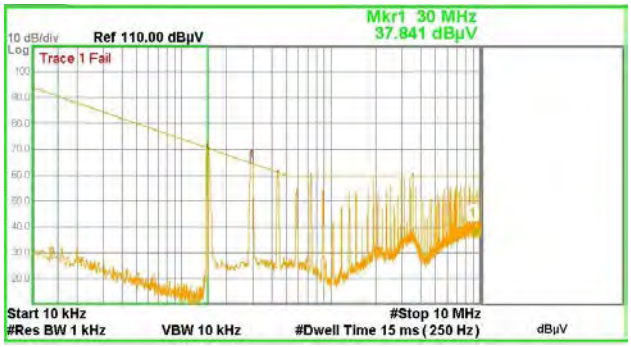


Figure 11. DM interference

4.3.2 CM component measurement

Equipment setup was changed according to Figure 7, power supply was on, load on EUT was om maximum, state CE102 was recalled and CM interference level was measured Figure 12.

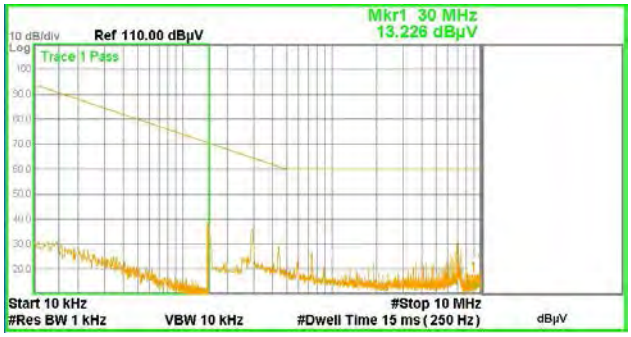


Figure 12. CM interference

4.3.3 Measurement data analysis and filter proposal

According to measurement results, on Figure 10, is shown that ambient interference is far below threshold curve for CE102 standard, so obtained measurement results are valid.

From Figure 11, is concluded that dominant interference, that EUT generate is DM, and EUT didn't passed test, its obvious that threshold curve for CE102 standard is exceeded, and in upper left corner of measurement data, Figure 11, is shown "Trace 1 Fail", so filter is needed to correct this type of interference.
On Figure 12, is shown that CM interference is significantly lower, for above 30 dB compared to DM interference, and this type is significantly below threshold curve for CE102, so this component type passed test, and for this component type, filter is not needed.

4.4 Filter proposal

For suppression of DM type of interference EMI filter is used, while for suppression of CM type of interference Choke filter is used.

As explained in privies chapter, for DM type of interference filter is needed.

The idea is to find EMI filter, with attenuation of DM type of interference, at least of 25dB or greater, according to Figure 11, and Figure 12. Also chosen EMI filter should be capable sustain maximum current capacity of EUT.

For purpose of explanation, filter FN409-6.5-02 is chosen. According to datasheet of this filter, filter have satisfying current capacity. Next, from datasheet, is analyzed attenuation chart Figure 13.

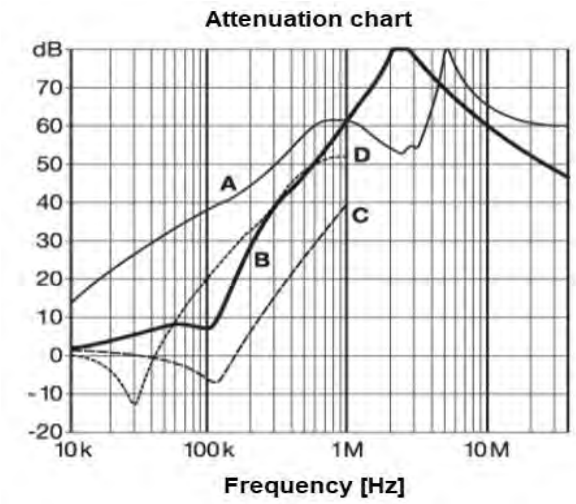


Figure 13. Typical Filter Attenuation [8]

Per CISPR 17 standard [9], each curve represents:
A = 50 Ω/50 Ω sym,
B = 50 Ω/50 Ω asym,
C = 0.1 Ω/100 Ω sym,
D = 100 Ω/0.1 Ω sym.
For calculating attenuation of DM type of interference, curve B is used.
and for calculating of CM type of interference, curve D is used. According to attenuation chart, Figure 13, measured value of DM and CM, Figure 11 and Figure 12, respectively, level of DM noise should be below 50dBµV, and level of CM noise should be at background noise level, after implementation of selected filter. Selected filter is installed on power line, between power supply and EUT, on the side of EUT. If the filter has ground connection, it connects to the chassis.

4.5 Verification

After filter implementation, as described in previous chapter, the tests, as shown on Figure 6 and Figure 7 were performed. Measurement results are showed on Figure 15 and Figure 16, respectively.

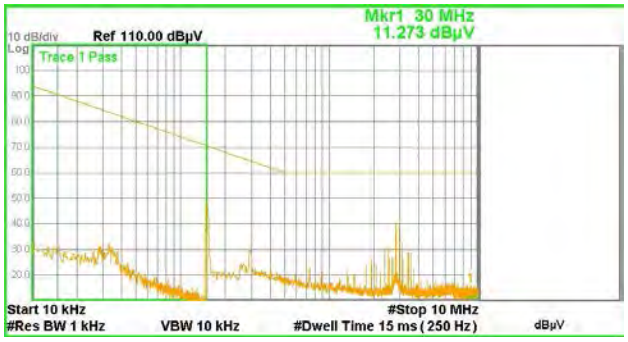


Figure 14. DM interference after filter implementation

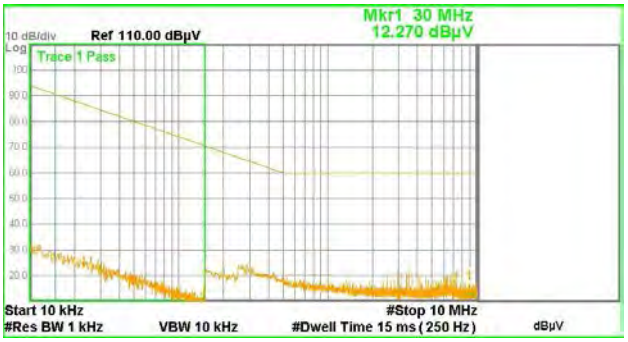


Figure 15. CM interference after filter implementation

Figure 14 and Figure 15, shows that analysis and calculation from previous chapter was correct.

Last test, that was carried out, for demonstration, that this methodology was correct, was test as shown on Figure 5, after a filter.

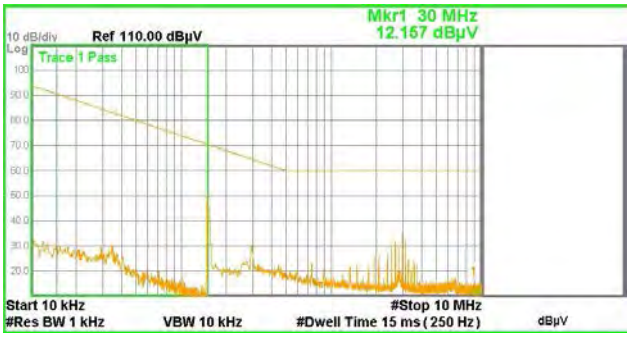


Figure 16. Conducted interference after filter implementation

On Figure 16, was shown that, also methodology proposed by MIL-STD-461, Figure 5, verified results. Only with expanded methodology, as proposed in this paper, and as showed on Figure 6 and Figure 7, is possible to manage each type of conducted interference separately.

5. CONCLUSION

This paper presents a measurement based methodology for the analysis of conducted electromagnetic interference, starting from the appropriate configuration of measurement equipment acquisition of conducted emission data, followed by spectral analysis and separation of the measured interference into DM and CM components. The obtained results are used to identify the dominant noise mode over investigated frequency range to determine required level of suppression. Further, based on required level of suppression and the dominant type interference, appropriate filter for implementation was selected. Finally, after the implementation of the filter, a final verification was performed to confirm that that the interference is within the limits defined by standard.

The future extension of this paper would include the design of filter, according to the data obtained by the methodology and measurement in order to optimize the filter for a specific device. Generic filters are usually not optimal, they cover wider frequency ranges, with certain suppression that could be much greater then is needed and because of that are certainly larger in size, heavier in mass and in higher price range. The designed filter would represent an optimal solution, would have required suppression, would be smaller in size, lighter and in lower price range.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] ANTONIJE R. ĐORĐEVIĆ, DRAGAN I. OLĆAN: *Ispitivanje Elektromagnetske Kompatobilnosti*, Univerzitet u Beogradu Elektrotehnički fakultet, Beograd, 2012.

[2] MIL-STD-461: https://quicksearch.dla.mil/qsDocDetails.aspx?ident_number=35789, last accessed 2026/08/09

[3] https://www.com-power.com/products/single-phase-lisns/li-3100, last accessed 2026/08/09

[4] https://www.tekbox.com/product/tblm2-lisn-mate/, last accessed 2026/08/09

[5] https://www.unitest.com/pdf/5990-7421EN.pdf, last accessed 2026/08/09

[6] https://www.keysight.com/us/en/product/909A/coaxial-termination-dc-18-ghz.html, last accessed 2026/08/09

[7] https://www.keysight.com/us/en/assets/7018-02033/data-sheets/5990-3453.pdf, last accessed 2026/08/09

[8] https://www.te.com/commerce/DocumentDelivery/DDEController?Action=srchrtrv&DocNm=FN409&DocType=Data%20Sheet&DocLang=English&DocFormat=pdf&PartCntxt=819398-SF, last accessed 2026/08/09

[9] https://www.normservis.cz/download/view/iec/info_cispr17{ed2.0}b.pdf, last accessed 2026/08/09

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

GEOSPATIAL ASSESSMENT OF THE DEVELOPMENT LEVEL OF LOCAL GOVERNMENT UNITS IN THE REPUBLIC OF SERBIA

ZORAN KRIČKOVIĆ
Military Geographical Institute – "General Stevan Bošković", Ministry of Defence of the Republic of Serbia, Mije Kovačevića 5, Belgrade, Serbia, zoran.krickovic@vs.rs

TIN LUKIĆ
University of Novi Sad, Faculty of Sciences, Department of Geography, Tourism and Hotel Management, Trg Dositeja Obradovića 3, Novi Sad, Serbia; University of Belgrade, Faculty of Geography, Studentski trg 3/3, Belgrade, Serbia; Ruđer Bošković School, Kneza Višeslava 17, Belgrade, Serbia, tin.lukic@dgt.uns.ac.rs

SAŠA BAKRAČ
Military Geographical Institute – "General Stevan Bošković", Ministry of Defence of the Republic of Serbia, Mije Kovačevića 5, Belgrade, Serbia, sasa.bakrac@vs.rs

VLADAN TADIĆ
Military Geographical Institute – "General Stevan Bošković", Ministry of Defence of the Republic of Serbia, Mije Kovačevića 5, Belgrade, Serbia, vladan.tadic@vs.rs

EMINA KRIČKOVIĆ
University of Belgrade, Faculty of Geography, Studentski trg 3/3, Belgrade, Serbia, emina.krickovic@gef.bg.ac.rs

Abstract: *Assessing the level of development of local government units (LGUs) represents one of the key instruments for understanding spatial disparities and supporting evidence-based regional policy. The methodological framework of this study integrates statistical and geospatial analyses to assess the level of development of LGUs in the Republic of Serbia. The analysis uses relevant socio-economic indicators. The Development Index of LGUs was calculated based on the following indicators: unemployment rate, income per capita, own revenue per capita, educational attainment, and population growth or decline rate. The highest development index was detected in Stara Pazova with 182.04 index points (i.p.), which means that the development of this LGU is 82.04 % above the republic average. Belgrade, as the largest city, is in the second place with 158.92 i.p., followed by Novi Sad (154.6 i.p.). Niš, as the third largest city in the Republic of Serbia, with 98.62 i.p. belongs to the second development group. In the fourth, the lowest development group, three LGUs detected development index lower than 50 % of the republic average – Rekovac (47.49 i.p.), Bojnik (49.19 i.p.), and Sjenica (49.62 i.p.). These findings highlight the need to reform the methodological framework, as well as the necessity of introducing new indicators that would more adequately reflect contemporary social, economic, and demographic challenges.*

Keywords: *Geospatial analysis, GIS, Local Government Units, development index, Republic of Serbia.*

1. INTRODUCTION

Assessing the level of development of local government units (LGUs) represents one of the key instruments for understanding spatial disparities and supporting evidence-based regional policy. Such assessments enable the identification of developed and underdeveloped areas, the allocation of development funds, and the implementation of measures aimed at achieving balanced regional growth.

The concept of territorial development is relatively recent and has evolved from earlier notions of regional and local development, as well as top-down and bottom-up approaches. Although these terms are often used interchangeably in practice, each represents a theoretical advancement over its predecessors. Local development, while frequently positioned within broader regional policy, remains insufficiently defined, yet is closely linked to key pillars such as capital and infrastructure policies, migration, and growth center strategies [1].

Development strategies have traditionally been framed through two contrasting approaches: top-down models based on industrialization and centralized planning, which often exacerbate inequalities, and bottom-up approaches that emphasize local resources, small-scale initiatives, and community participation [2].

More recently, the concept of territorial development has emerged as an integrative framework that encompasses local and regional development while incorporating notions such as place, clusters, and ecosystems [3]. Contemporary perspectives further highlight the role of both production processes and governance structures as

Corresponding author: Z. Kričković;
E-mail: zoran.krickovic@vs.rs
DOI: 10.5937/OTEH26098K

key drivers of territorial development, alongside broader forms of innovation, including social and institutional change, often shaped through both cooperation and conflict [4].

Regional disparities in Serbia remain a significant development challenge [5]. Although a legal and institutional framework exists, anchored in the Constitution and the Law on Regional Development [6], its implementation is constrained by outdated strategic documents and limited policy updates. The classification and measurement of LGU development levels represent a key instrument of regional policy [7], yet no revised regulatory framework has been adopted in recent years. This underscores the need for updated methodologies and policies to support more balanced and sustainable territorial development.

In Serbia, as in many other countries of Central and Eastern Europe, significant regional inequalities persist, both in terms of economic potential and quality of life [8]. The use of geospatial analysis in this context enables a more detailed examination of spatial patterns of development, revealing the spatial concentration of more or less developed municipalities. By applying GIS tools and statistical methods, it becomes possible to visualize the spatial distribution of indicators of socio-economic development and to detect LGUs that share similar development characteristics. This approach contributes not only to a better understanding of spatial disparities, but also to the formulation of more targeted and efficient development policies.

The methodological framework of this study is based on the integration of statistical and geospatial analyses aimed at assessing the level of development of LGUs in the Republic of Serbia. The analysis includes the use of relevant socio-economic indicators that reflect the demographic, economic, and infrastructural characteristics of municipalities and cities.

2. METHODOLOGY

The study area encompasses the entire territory of the Republic of Serbia, including its LGUs. According to the data of the Statistical Office of the Republic of Serbia (SORS), there are 197 municipalities, city municipalities, and cities that constitute the LGUs [9], [10], [11]. Out of the total number of LGUs, 29 are located in the region of Kosovo and Metohija, for which statistical data are not available in the official databases of SORS. Consequently, these LGUs were excluded from this research.

Spatial data used in this analysis were obtained from the open data platform of the Republic Geodetic Authority (RGA) [12]. As the results of this study are presented at the level of cities and municipalities, it was necessary to adjust the available RGA datasets accordingly. Figure 1 presents an overview of the research workflow applied in this study.

In this process, the city municipalities of Belgrade, Niš, Vranje, Požarevac, and Užice were merged into single city polygons, thereby forming 145 LGUs analysed in total across the Republic of Serbia. The spatial processing and aggregation were conducted in ArcGIS Pro 3.0. During the research, Microsoft Excel was used for the initial processing and cleaning of open datasets obtained from SORS. This step included the identification, extraction, and preparation of relevant data to be used for further analysis. According to the Official Gazette [13], the Development Index of LGUs is calculated based on the following indicators:

– Unemployment rate, calculated as the ratio between the number of unemployed persons and the sum of employed and unemployed persons within an LGU;
– Income per capita, representing the ratio of the total taxable income earned in a given calendar year (by the place of residence of the income recipient) and the total population of the LGU;
– Own revenue per capita, calculated as the ratio between the total realized own revenues of an LGU and its total population;
– Educational attainment, expressed as the share of the working-age population with completed secondary, post-secondary, or higher education in the total working-age population (according to the most recent census); and
– Population growth or decline rate, calculated as the ratio between the population size in the observed year and that recorded in the 2002 census at the LGU level.

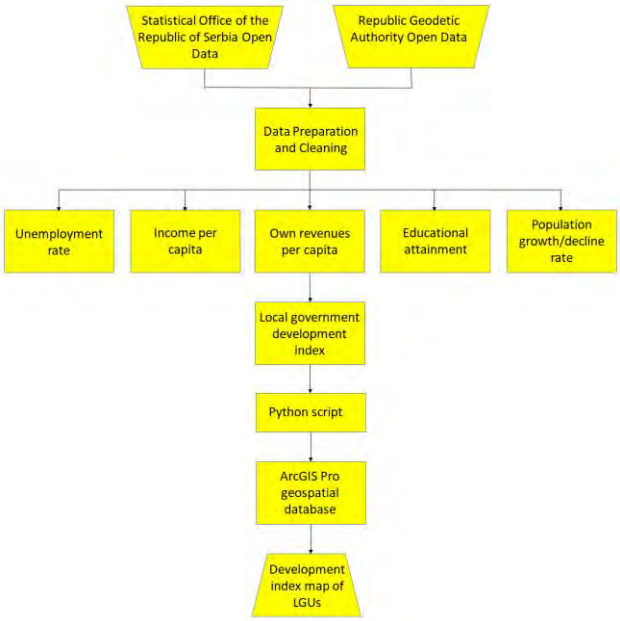


Figure 1. Schematic process of the research.

After calculating all indicators that reflect the level of development of a given territory or region, they were used to prepare an updated Unified Development List of LGUs in Serbia. According to the Official Gazette [13], each indicator is assigned a specific weight coefficient within the composite Development Index:

– Unemployment rate – 30%,
– Income per capita – 25%,
– Own revenue per capita – 10%,

– Educational attainment – 15%,

– Population growth or decline rate – 20%.

For the calculation of unemployment, income, and own revenue per capita, data for the three years preceding the year of calculation were used. In this study, that period covered 2022-2024. Additionally, a correction factor of -10 points was applied for LGUs in the fourth development group, whose working-age population density is lower than 30 persons per square kilometre. The calculation of the unemployment rate indicator was performed using the following equation:

$$I_{j,i} = \frac{R_i}{P_{j,i}} \times 100 \quad (1)$$

where I is the relative index (the national average equals 100), j represents the LGU, i the indicator, R the national average, and P the value of the indicator for the given LGU. For other indicators, the inverse proportion was applied, as higher values indicate better performance. According to these proportions, for the unemployment rate, values greater than 100 indicate results above the national average, while values below 100 indicate below-average performance. For all other indicators, higher values denote better results relative to the national average.

A review of the STAT database of the SORS and the publications Municipalities and Regions in the Republic of Serbia ([9], [10], [11]) revealed that data on income per capita at the municipal level are not available. Following direct communication with the Communication and Information Department of the SORS via email, it was confirmed that the SORS does not possess such data, but only data on the average wages of employees by municipality of residence and on household income at the regional level. The use of regional-level income data was considered inappropriate, as disaggregation to the level of LGUs would likely result in significant inaccuracies. Therefore, this study employs data on the average wages of employees by municipality and the number of employees as indicators of the economic status of the population, enabling an approximate estimation of income per capita. The Tax Administration of the Republic of Serbia undoubtedly collects such data, as clearly outlined in its Tax Guide [14], which specifies the categories of taxable income, including wages, taxable income from self-employment, copyright and related rights and industrial property rights, real estate, the rental of movable property, from the provision of hospitality services, taxable income of athletes and sports professionals, seafarers, and other taxable income [14].

In this context, foreign income refers to income earned and taxed in another country by taxpayers who are residents of the Republic of Serbia. Therefore, average wages represent only one of nine income categories that would be relevant for calculating the income per capita indicator. Consequently, this indicator, when estimated solely on the basis of average wages at the level of LGUs, does not fully reflect the actual socio-economic conditions of the population within a given territory or region. However, due to the lack of publicly available data, it is not possible to calculate income per capita in a precise and comprehensive manner. Data for the next indicator, own-source revenue per capita, were used on current revenues data from the SORS. The reason is the same as for the previous indicator – official publications of the SORS do not provide data on own-source revenues of LGUs, but only on current revenues.

The level of education, as the following indicator of territorial or regional development is defined as the ratio between the working-age population with completed secondary, post-secondary, or higher education and the total working-age population. During the analysis of this indicator certain inconsistencies were identified. Specifically, if the working-age population is assumed to comprise individuals aged 15 to 64, and the number of individuals with the aforementioned levels of education is aggregated based on the 2022 census data [15], the resulting share of the education level exceeds 100% in some LGUs. This clearly represents an unrealistic depiction of the actual educational structure of the population. It is established that the publications of the SORS do not present the number of working-age population by educational structure, but rather only data for the population aged 15 and over. For this reason, the total population aged 15 and above was used for the purposes of this research, in order to ensure that differences in educational structure remain positive and suitable for further analysis.

Based on the SORS data on 2002 census [15] and the projection of the population number at the end of 2024 [11], the population growth or decline rate by LGU is calculated. After applying the assigned weights, the final Development Index (N) for each LGU (j) was calculated using the following equation:

$$N_j = \omega_1 I_{j1} + \omega_2 I_{j2} + \omega_3 I_{j3} + \omega_4 I_{j4} + \omega_5 I_{j5} \quad (2)$$

The correction factor mentioned above was then applied to the final index values for the fourth development group. As data on Kosovo and Metohija region for some indicators were not available, therefore the development index for this region couldn't be calculated.

As a link between the data and the obtained results in MS Excel, a script written in the Python programming language was used. The open-source software PyCharm, a popular Integrated Development Environment (IDE) for Python, served as the main development platform. Today, Python offers a large number of libraries applicable to various fields. In this research, in addition to the core libraries, two additional libraries were used: openpyxl and arcpy.

The openpyxl library is used for working with Excel files in the .xlsx format. It enables reading, creating, editing, and saving Excel files without requiring MS Excel itself. Its key functionalities include reading and writing data to cells, formatting (colours, fonts, styles), working with multiple sheets, creating formulas, charts, and tables, as well as manipulating rows and columns [16].

The second library, arcpy, was developed by Esri for use with ArcGIS Pro (and formerly ArcMap). It allows for the

automation of GIS workflows – including spatial analysis, geospatial data processing, management of raster and vector data, and the administration of geodatabases. The key advantage of this library lies in its ability to automate spatial analyses [17]. Arcpy serves as the foundation for the integration of ArcGIS Pro within the Python environment and provides a vital link between various data sources and geospatial databases.

3. RESULTS AND DISCUSSION

According to the Official Gazette [13], out of the five regions that constitute the second level of the Nomenclature of Statistical Territorial Units (NUTS 2) of the Republic of Serbia, two (Belgrade and Vojvodina regions) are classified as developed, while the remaining three, located in the southern part of the country (Šumadija and Western Serbia, Southern and Eastern Serbia, and Kosovo and Metohija regions), are considered underdeveloped. It is evident that the northern part of the Republic is more developed than the southern part. Of the total territory of the Republic of Serbia, 28% is classified as developed, whereas the remaining 72% is underdeveloped. Thus, the developed area accounts for less than one-third of the country's total territory. The geographical distribution of development clearly reflects the inequality in the territorial development of the Republic of Serbia.

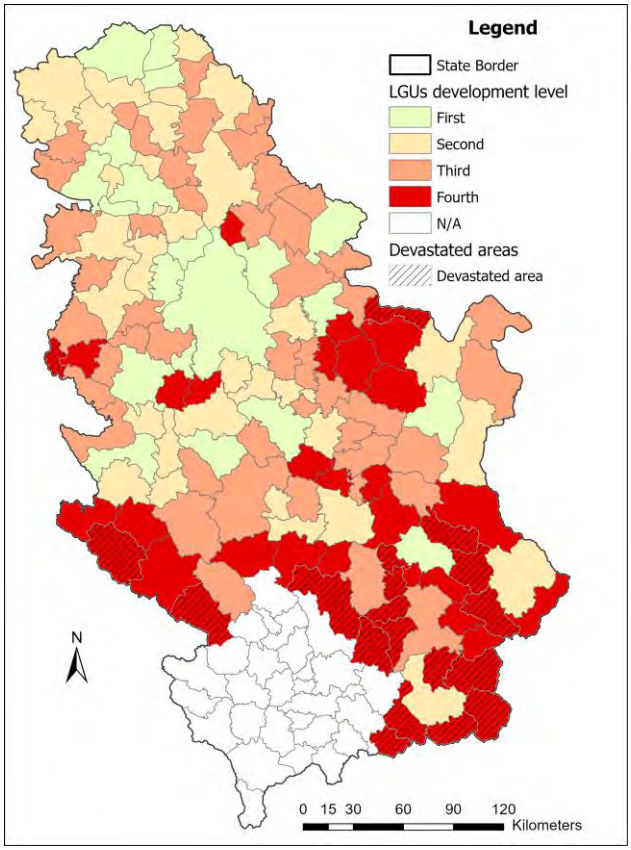


Figure 2. Spatial LGUs development according the Unified List of Development Levels of Regions and LGUs for 2014 [13].

Spatial distribution of the development of cities and municipalities within the territory of the Republic of Serbia is compiled according to the Regulation on the Determination of a Unified List of Development Levels of Regions and LGUs for 2014 (Figure 2). According to this overview, out of a total of 145 cities and municipalities, 20 belong to the first, most developed group, 34 to the second, 47 to the third, and 44 to the least developed, fourth group. Of these 44 LGUs from the least developed group, 19 are classified as devastated areas. Only 13.78% of LGUs have a development level above the national average, while nearly one-third (30.35%) have a development level below 60% of the national average. Devastated areas account for 13.1% of the total number of LGUs in the Republic of Serbia. It is evident that the least developed LGUs and the most devastated areas are located in the southern part of the Republic of Serbia, indicating significant regional disparities in territorial development.

As the first indicator of the development index, the unemployment rate shows a clear spatial disparity in the unemployment rate across the territory of the Republic of Serbia can be observed, with the highest levels recorded in the southwestern and southeastern parts of the country. The highest unemployment rates, exceeding 50%, were registered in Tutin (64.39%) and Novi Pazar (53.05%), while the lowest rates were observed in Stara Pazova (4.18%), Žagubica (6.53%), and Bački Petrovac (6.9%). Belgrade, as the capital city, follows closely with an unemployment rate of 7.01%.

The second indicator, Income per capita, calculated according to the method described above, also clearly highlights the higher level of development in major urban centres such as Belgrade and Novi Sad. An exception is observed in Bor, where the highest income per capita was recorded. In contrast, LGUs in the south-western and southern parts of Serbia, including Novi Pazar, Sjenica, Tutin, Preševo, and Bujanovac, exhibit the lowest levels of this indicator.

Due to the use of the aforementioned data, the results are somewhat overestimated for the third indicator – own revenue per capita, as the current revenues exceed own-source revenues. This indicator clearly reveals differences in relation to LGUs with smaller populations, where the highest own-source revenues per capita were recorded, such as in Čajetina (194.37 thousand RSD), Crna Trava (169.38 thousand RSD), Trgovište (148.12 thousand RSD), and Bor (142.5 thousand RSD), which thus represent the LGUs with the highest values of this indicator.

In contrast, LGUs with larger populations exhibit lower own-source revenues per capita, as observed in Bujanovac (27.14 thousand RSD), Preševo (28.15 thousand RSD), and Tutin (30.58 thousand RSD). As illustrated on the map, the majority of LGUs fall within the lowest category of own-source revenues (70 LGUs), followed by 56 LGUs in the second group and 19 LGUs in the third group, which includes the City of Belgrade as the largest urban center in the Republic of Serbia.

The aforementioned inconsistencies in the calculation of the following indicator of the development index – educational attainment, can be illustrated by the example

of the capital city, Belgrade. The SORS 2022 census data of the population between 15-64 years old for this city is 1,086,113. When the number of the population older than 15 years with completed secondary, post-secondary, or higher education is added, the result is 1,235,175 population. This is negative difference and this calculation is impossible to use. The uneven development of LGUs is clearly visible, with more populous cities generally exhibiting higher educational attainment, such as Novi Sad (86.68 %), Belgrade (86.55 %) and Niš (84.80 %). Unlike these, cities in Braničevo District have the lowest educational attainment – Malo Crniće (38.55 %), Žabari (43.38 %), and Petrovac na Mlavi (44.7 %).

The last indicator – population growth shows that population growth was detected in only four LGUs – Novi Pazar (126.75 %), Novi Sad (124.59 %), Tutin (111.13 %), and Belgrade (106.72 %). Unlike these, in the remaining 97% of LGUs population decline was detected, with the highest percentage detected in Crna Trava, with only 37.22 % of the population from 2002.

After all indicators calculated, the development index is calculated according earlier mentioned methodology. The geospatial distribution of LGUs according to the calculated Development Index is shown in Figure 3. It is clear the development progress of the LGUs, as in the lowest group are 13 LGUs, in the second, most number of the LGUs was detected – 62, in the third 42 and in the first group, in which development is above republic average, are 29 LGUs. Also, uneven territorial development of the Republic of Serbia is visible.

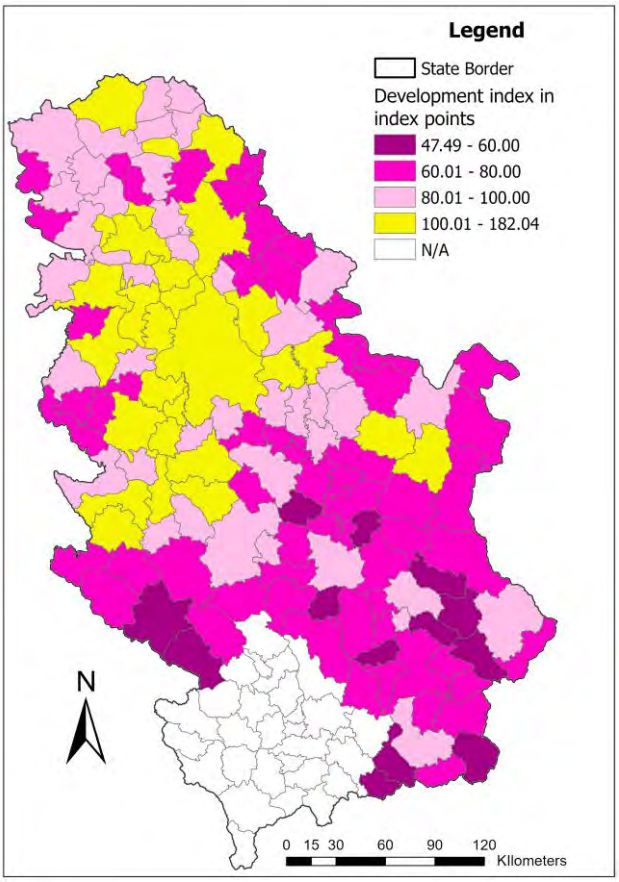


Figure 3. Development index for 2024 year according to publicly available SORS data.

The highest development index was detected in Stara Pazova with 182.04 index points (i.p.), which means that the development of this LGU is 82.04 % above republic average. Belgrade, as the largest city is in the second place with 158.92 i.p. followed by Novi Sad (154.6 i.p.). Niš, as the third largest city in the Republic of Serbia, with 98.62 i.p. belongs to the second development group. In the fourth, the lowest development group, three LGUs detected development index lower then 50 % of the republic average – Rekovac (47.49 i.p.), Bojnik (49.19 i.p.), and Sjenica (49.62 i.p.).

If these results are visually compared with the new methodologies proposed in [7], in which the authors identified two variants that, in terms of all structural performance indicators, provide the most realistic representation of the level of development of municipalities in the Republic of Serbia [7], it can be concluded that they yield broadly similar results with regard to the territorial distribution of the development index, although individual outcomes differ significantly. This is particularly evident in the case of Žagubica, which, according to this study, belongs to the first group of developed LGUs (108.77 i.p.), whereas in [7] it is classified within the fourth group, comprising devastated areas. Other LGUs do not exhibit such pronounced differences in the development index between this study and the results presented in [7].

In the context of the Republic of Serbia, the analysis indicates significant regional disparities. Although a well-developed legal and institutional framework for monitoring the level of development of LGUs exists [6], and [13], the continued application of a unified classification list from 2014 points to issues related to insufficient updating and the limitations of existing mechanisms, as also discussed in [7]. The results of this study, based on the most recent statistical data, confirm pronounced inequalities, particularly between the northern and southern parts of the country, furthermore that more than one-third of LGUs fall into the category of underdeveloped or devastated areas.

The assessment of local self-government development in Serbia, based on the official development methodology and the most recent statistical indicators, provides valuable insights into the spatial distribution of demographic, economic, infrastructural, and public service capacities across municipalities. The results of this research can significantly support the identification and prioritization of safe helicopter landing zones within the Project in the Military Geographical Institute “The use of geospatial data within a multicriteria analysis for the identification of safe helicopter landing zones in the territory of the Republic of Serbia”, as these development indicators may provide useful contextual information for future multicriteria analyses of transport accessibility, healthcare infrastructure, emergency response capacity, and public-space availability. Integrating development indicators with geospatial criteria may therefore contribute to a more efficient and equitable planning of helicopter landing networks, particularly in areas requiring rapid access to emergency medical and disaster response services.

Furthermore, these findings highlight the need to reform the methodological framework, as well as the necessity of introducing new indicators that would more adequately reflect contemporary social, economic, and demographic challenges. Furthermore, strengthening the role of local communities in the formulation of development policies is essential, in order to enable the inclusion of specific local resources and potentials.

4. CONCLUSION

The comparative analysis of the development index has shown that Serbia is still characterized by a high degree of inequality in the level of regional and LGU development, which significantly limits the possibilities for balanced and sustainable development. Although the existing institutional and legal mechanisms are well established in theory, in practice they are not sufficiently updated or adequately implemented, as reflected in the continued use of the 2014 Regulation. The data obtained in this study indicate the need to develop a new, up-to-date unified classification of LGU development levels, which would provide a more realistic representation of the current situation and serve as a basis for more effective regional policy planning.

In the future, it is necessary to design a model that would align the interests of local communities with national development strategies and introduce modern indicators that reflect the complexity of territorial development, as also proposed in the Study of the Republic of Serbia and its collaborators. Only through the integration of economic, social, and demographic dimensions, together with the active participation of all stakeholders, can a more equitable distribution of resources and a more balanced territorial development of the Republic of Serbia be achieved.

Acknowledgement

This study was supported by the project of the Ministry of Defence of the Republic of Serbia and the Serbian Armed Forces (Project No. 1.18-2026). Also, this study was supported by the Program of Cooperation with the Serbian Scientific Diaspora- Joint Research Projects – DIASPORA 2023, under the project LAMINATION (Project No: 17807). Furthermore, this study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grant No 451-03-33/2026-03/200125 & 451-03-34/2026-03/200125 and 451-03-33/2026-03/200091).

References

[1] W. J. Coffey and M. Polèse, “Local development: Conceptual bases and policy implications,” Regional Studies , vol. 19, no. 2, p. 85–93, 1985.

[2] W. Stöhr and D. R. F. Taylor, “Development from above or below? The dialectics of regional planning in developing countries,” WU Vienna University of Economics and Business, Vienna, vol. 2: Regional Development Alternatives: International Perspectives, pp. 9-27, 1981.

[3] A. Torre, “Territorial development: towards a dynamic and innovative understanding,” Regional Studies, vol. 59, no. 1, 2025.

[4] A. Torre, “Contribution to the theory of territorial development: a territorial innovations approach,” Regional Studies, vol. 59, no. 1, 2025.

[5] Petrović J, Radukić S, Radović M. Multicriteria analysis of the level of demographic development of regions in Serbia. Poslovna ekonomija. 2017;12(2). Serbian.

[6] Official Gazette of the Republic of Serbia. Law on Regional Development. Official Gazette of the Republic of Serbia. 2009;51/2009.

[7] Šećerov V, Filipović D, Čokorilo N, Tontić S, Šantić D, Bjelobrk G, Vračarević B, Jeftić M, Duškov Lj, Protić B. New model for measuring the level of development of local self-government units in the Republic of Serbia. Belgrade: Republic of Serbia, Cabinet of the Minister for Balanced Regional Development; University of Belgrade, Faculty of Geography; 2023. Serbian.

[8] Heyns B. Emerging inequalities in Central and Eastern Europe. Annu Rev Sociol. 2005;31:163-197. doi:10.1146/annurev.soc.30.012703.110637

[9] Statistical Office of the Republic of Serbia. Municipalities and regions in the Republic of Serbia. Belgrade: Statistical Office of the Republic of Serbia; 2022.

[10] Statistical Office of the Republic of Serbia. Municipalities and regions in the Republic of Serbia. Belgrade: Statistical Office of the Republic of Serbia; 2023.

[11] Statistical Office of the Republic of Serbia. Municipalities and regions in the Republic of Serbia. Belgrade: Statistical Office of the Republic of Serbia; 2024.

[12] Republic Geodetic Authority. GeoSerbia e-Services: Data download [Internet]. Available from: https://download.geosrbija.rs/download/login [cited 2025 Aug 8].

[13] Official Gazette of the Republic of Serbia. Decree on establishing the unified list of development levels of regions and local self-government units for 2014. Belgrade: Official Gazette of the Republic of Serbia; 2015.

[14] Republic of Serbia, Ministry of Finance, Tax Administration. Tax Informer: Annual personal income tax [Internet]. April 2025. Available from: https://purs.gov.rs/upload/media/2025/4/11/759560/Godisnji_porez_na_dohodak_gradjana.pdf [cited 2025 Sep].

[15] Statistical Office of the Republic of Serbia. 2022 Census of Population, Households and Dwellings: Data by municipalities and cities. Belgrade: Statistical Office of the Republic of Serbia; 2023.

[16] OpenPyXL Documentation, “OpenPyXL documentation. Read the Docs,” [Online]. Available: https://openpyxl.readthedocs.io/. [Accessed 2025].

[17] Esri, “What is ArcPy? In ArcGIS Pro documentation,” Esri, [Online]. Available: https://pro.arcgis.com/en/pro-app/arcpy/get-started/what-is-arcpy-.htm. [Accessed 2025].

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AUTOMATING DIGITAL MAP PRODUCTION PROCESSES USING PYTHON GIS ALGORITHMS

MARKO SIMIĆ
Military Geographical Institute – "General Stevan Bošković", Belgrade, marko.simic@vs.rs

SAŠA BAKRAČ
Military Geographical Institute – "General Stevan Bošković", Belgrade, sasa.bakrac@vs.rs

ĐORĐE ĐERMANOVIĆ
Military Geographical Institute – "General Stevan Bošković", Belgrade, djordjedjermanovic99@gmail.com

MATIJA MILIĆ
University of Niš, Faculty of Science and Mathematics, Niš, matija.milic@pmf.edu.rs

SLAVIŠA TATOMIROVIĆ
Faculty of Diplomacy and Security, Union University – "Nikola Tesla", Belgrade, Slavisat04@gmail.com

Abstract: *The development of Python-based algorithms has enabled the automation of selected processes in digital topographic map production at the Military Geographical Institute – "General Stevan Bošković" (MGI). In this study, automation is presented as a broader strategy for improving efficiency, consistency, and sustainability in GIS-supported map production workflows. Algorithm development was moved from ArcMap to ArcGIS Pro in order to support the transition from Python 2 to Python 3 and to provide a more modern, maintainable, and extensible software environment. Only algorithms that depend on specific ArcMap functionality, such as legacy representation features, were retained in the ArcMap environment. Implemented as native GIS tools, the algorithms incorporate institutional cartographic rules into modular geoprocessing procedures that can be used by both developers and GIS operators. As an illustrative example, the paper highlights the automation of content alignment along adjacent map sheet boundaries in ArcGIS Pro, a traditionally labor-intensive task requiring detailed visual inspection and manual correction. Particular attention is given to the Python-based processing pipeline and to the formulation of the functions used for geometry refinement, consistency control, and operator-guided validation of remaining discrepancies. The migration to ArcGIS Pro also improved compatibility with current GIS technologies and established a stronger foundation for future upgrades, including adaptation to different cartographic products and scales. The results show that Python GIS algorithms can significantly reduce manual intervention, improve workflow consistency, and provide a transferable framework for the modernization of digital map production processes.*

Keywords: *GIS, Python, Geoprocessing Automation, Digital Topographic Maps.*

1. INTRODUCTION

Digital topographic map production is no longer limited to the visual design and reproduction of cartographic content. In modern production systems, topographic maps are created from complex geospatial databases in which geometric accuracy, attribute consistency, topological correctness, and repeatability of procedures are equally important. This has increased the importance of automated and semi-automated GIS workflows, especially in institutions responsible for systematic map production [1].

At the same time, the growing volume of geospatial data has made purely manual cartographic procedures increasingly difficult to maintain. Tasks that were traditionally performed through visual inspection and manual editing often require repeated execution on large numbers of features. Such procedures are time-consuming and may lead to inconsistent results, since repetitive manual editing increases operator fatigue and the possibility of overlooked error. For this reason, automation has become an important element of digital cartography, contributing to reduced processing time, standardized production rules, and improved quality of the final geospatial database.

Python has become one of the most widely used programming languages for GIS automation because it provides a flexible environment for processing spatial data and building customized geoprocessing workflows. In GIS applications, Python is commonly used for data conversion, spatial analysis, topology control, batch processing, and the integration of multiple geoprocessing operations into repeatable procedures. Through libraries such as ArcPy, PyQGIS, GDAL/OGR, GeoPandas, Shapely, and PyProj, Python enables the development of tools for practical production procedures in which institutional production guidelines are directly

Corresponding author: M. Simić;
E-mail: marko.simic@vs.rs
DOI: 10.5937/OTEH26099S

implemented [2, 3].

The role of Python in geospatial applications is not limited to conventional geoprocessing automation. Python-based workflows are also increasingly employed for the integration of GIS and remote-sensing data, automated spatial analysis, and machine- and deep-learning applications [4–6].

Within the ArcGIS environment, Python has an important role because it allows users to transform individual geoprocessing operations into structured tools that can be used directly from the program interface. Depending on the complexity of the task, Python code can be implemented as standalone scripts, script tools, or Python toolboxes. These approaches make it possible to automate repetitive procedures, define user parameters, control input data, and integrate developed tools into existing production workflows [7, 8]. Therefore, Python-based automation represents a practical bridge between software development and everyday cartographic production.

In the production of digital topographic maps at the Military Geographical Institute – "General Stevan Bošković" (MGI), Python-based algorithms have been developed to support selected cartographic and geoprocessing procedures. These algorithms are designed to incorporate institutional cartographic rules into repeatable GIS tools, while still allowing operator supervision where full automation is not appropriate. This is particularly important in topographic map production, where the final result must satisfy both geometric and cartographic requirements. Earlier automation procedures were primarily developed in the ArcMap environment and were based on Python 2. Although these tools provided practical benefits in production, the long-term sustainability of such workflows depends on their modernization and adaptation to current GIS technologies [9]. The transition from ArcMap to ArcGIS Pro and from Python 2 to Python 3 requires the revision of existing algorithms, adaptation of geoprocessing procedures, restructuring of code, and evaluation of which functions can be transferred to the new environment and which must remain connected to legacy ArcMap functionality.

An analysis of Python GIS algorithms is conducted within the broader framework of modernizing digital topographic map production. The focus is placed on the organization of automated workflows, the migration of algorithm development toward ArcGIS Pro, and the use of modular geoprocessing procedures. Content alignment along adjacent map sheet boundaries is used as an example of this approach because it represents a complex and traditionally labor-intensive procedure that combines geometry refinement, consistency control, and operator-guided validation. More broadly, the migration to Python 3 and ArcGIS Pro establishes a sustainable framework for the automation of additional production tasks, such as cartographic generalization in the derivation of smaller-scale maps from the 1:25,000 digital topographic map (DTM25).

2. MIGRATION FROM PYTHON 2 TO PYTHON 3

The migration from Python 2 to Python 3 represents an important step in the modernization of GIS-supported cartographic production. Earlier automation procedures were developed in the ArcMap environment which relied on Python 2. Although these tools were functional and useful in production, their long-term use is limited by software obsolescence, reduced compatibility with modern libraries, and the gradual transition of GIS platforms toward Python 3 [10]. ArcGIS Pro uses Python 3 and represents the current development environment for Python-based GIS automation in the Esri ecosystem.

This migration is not only a technical update of programming syntax, but also a transition toward a more sustainable production framework. Maintaining algorithms in an outdated software environment can create several practical problems, including limited support for new geoprocessing tools, difficulties in integrating updated Python libraries, reduced compatibility with new data formats, and increased dependence on legacy workflows. Since ArcGIS Desktop, including ArcMap, has reached the end of its product life cycle, migration to ArcGIS Pro is also necessary for ensuring the continuity of institutional GIS workflows.

One of the main advantages of migration to ArcGIS Pro is the use of a modern Python 3 environment managed through Conda. This enables better control of Python packages, easier management of project-specific environments, and access to updated scientific and geospatial libraries. This environment contains ArcPy and additional Python libraries used for geoprocessing, data analysis, and automation [11].

However, migration from Python 2 to Python 3 also requires careful revision of existing algorithms. Certain changes in syntax, string handling, dictionary behavior, module organization, and library compatibility can require modification of older code. Automated conversion tools can assist in this process, but they do not fully replace manual inspection and testing. For this reason, migration represents a controlled redesign of the workflow rather than a simple conversion of scripts from one Python version to another [12].

In the context of digital topographic map production at the MGI, this process involved analysing the existing ArcMap-based algorithms and determining how their logic could be transferred to the ArcGIS Pro environment. For the workflow presented in this study, complete migration was possible because the procedure did not depend on legacy cartographic representation mechanisms specific to ArcMap. The algorithms were therefore redesigned for Python 3 and ArcGIS Pro while preserving the original cartographic logic. In this way, an established production task was transferred into a modern software environment, improving maintainability, compatibility with current GIS technologies, and the potential for future expansion.

By adapting the tools to a modern GIS environment, the code becomes easier to update, document, test, and extend. This is especially important in institutional cartographic

production, where tools are not created for one-time use, but must remain operational across multiple production cycles. A modernized Python 3 workflow also makes it easier to introduce additional automated procedures, connect different processing stages, and gradually reduce repetitive manual work.

3. MATERIALS AND METHODS

DTM25 production at the MGI includes several interconnected phases, such as editorial preparation, aerial photogrammetric survey, photogrammetric mapping, field verification and completion, cartographic processing, and reproduction [9]. This paper focuses on the cartographic processing phase, which includes numerous repetitive geoprocessing operations that require consistency, precision, and repeatability across different production tasks.

The selected production procedure used for methodological demonstration is the alignment of content along adjacent DTM25 map sheet boundaries. This procedure was selected because the alignment of adjacent map sheets requires a combination of interrelated operations, including the detection of features near sheet boundaries, geometric refinement of line and polygon features, control of topological continuity between neighbouring sheets, and identification of remaining discrepancies for operator review. In manual workflows, this task requires detailed visual inspection of all feature vectors located near the map frame. Small geometric offsets, gaps, or overlaps may remain unnoticed at the standard display scale, especially in areas with dense content such as contour lines, hydrography, or communication networks.

The automated workflow is organized around three production scenarios. The first scenario refers to the case in which the adjacent sheet has already been completed, so the active sheet is adjusted according to the finalized geometry of the neighbouring sheet. The second scenario refers to the simultaneous processing of two adjacent sheets, where both sheets are adjusted in a coordinated manner. The third scenario refers to the preparation of an active sheet when the neighbouring sheet has not yet entered the cartographic processing phase. These scenarios correspond to unidirectional, bidirectional, and preparatory alignment procedures.

The input data consist of DTM25 feature classes stored in production geodatabases, together with the sheet index polygon layer used to define the boundaries of adjacent map sheets. The procedure processes line and polygon feature classes separately because these geometry types require different correction logic. Line features include contour lines, roads, railways, flowing water, communication objects, and other linear cartographic elements. Polygon features include vegetation areas, water bodies, marshland, and other surface elements (Figure 1.).

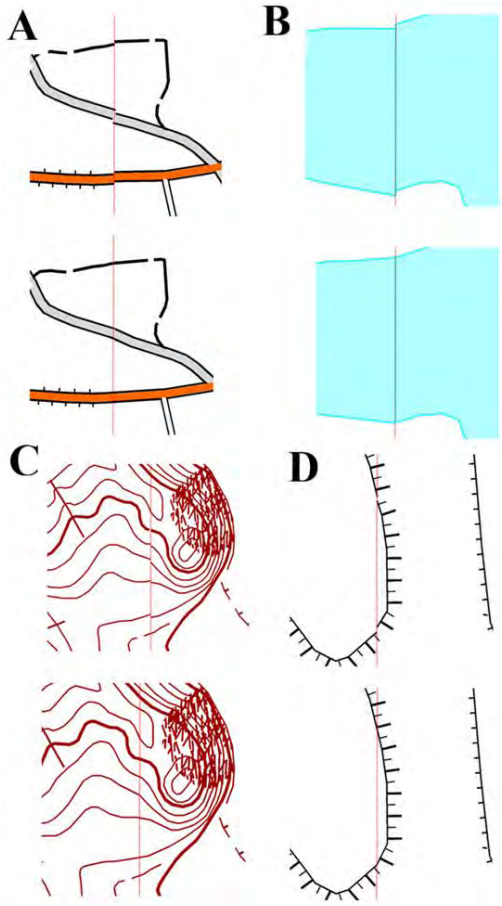


Figure 1. Examples of communication networks (A), hydrology (B), contour lines (C), and relief (D) at TK25 sheet boundaries (red line) before (top image) and after (bottom image) automatic vector processing

Point features are used as control outputs, marking locations where automatic processing detected unresolved or potentially problematic geometry (Figure 2.).

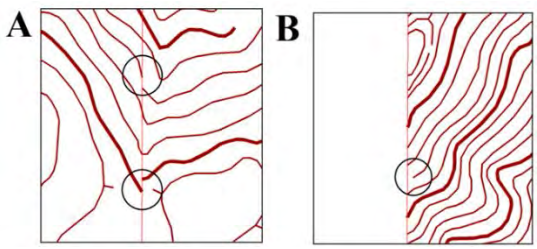


Figure 2. Point indicators (displayed as circles) marking unresolved contour-line geometry in bidirectional alignment (A) and preparatory alignment (B)

The workflow was implemented in ArcGIS Pro using Python 3 and ArcPy geoprocessing functions. The tools were designed as modular GIS procedures that can be executed from the ArcGIS Pro interface, allowing operators to use them without direct interaction with the source code. Before running the procedure, the operator defines the relevant input geodatabases, the map sheet pair, the feature classes to be processed, and the tolerance values for geometry correction. Repetitive and rule-based operations are performed automatically, while the operator remains responsible for parameter selection and validation of marked discrepancies making the workflow semi-

automated.

The general processing sequence consists of four main steps. First, user-defined parameters are read and assigned to internal variables. Second, the polygons of the selected DTM25 sheet pair are extracted from the sheet index layer and used as spatial references for selecting features along the boundary zone. Third, separate processing objects are initialized for line and polygon feature classes. Fourth, the tool performs hierarchical processing. It first iterates through each selected sheet pair and then through the selected feature classes within that pair. This structure allows the same logic to be applied to multiple sheets and multiple cartographic layers without rewriting the procedure for each individual case.

To make the processing logic more explicit, the workflow can be represented as a Python pipeline composed of reusable geoprocessing functions. This functional representation separates the algorithm into elementary operations. Formulating the process in this manner is useful because it shows the order of execution, and how the same functions can be reused in different production scenarios.

If F_L is defined as the function for processing line features, F_G as the function for processing polygon features, and P_L as the function for creating point features that indicate changes or unresolved discrepancies, the procedures can be represented as follows.

Active and Passive Replic

For line features, the active sheet is adjusted according to the passive, already finalized neighbouring sheet:

$$F_L(e) = (S_2 \circ L \circ S \circ I_2)(e) \quad (1)$$

$$P_L(e) = \big(P \circ M(I, I_2)\big)(e) \quad (2)$$

For polygon features, the algorithm first selects the relevant polygon elements on both sheets and then applies snapping and densification along the boundary zone:

$$G(e) = (L_A \circ L_1)(e) \quad (3)$$

$$G_2(e) = (L_A \circ L_2)(e) \quad (4)$$

Two Active Replicas

In the case of two active replicas, the alignment is performed in two phases because both neighbouring sheets are still in production.

For line features, the first phase prepares and adjusts selected features from both sheets:

$$B(e) = (I \circ L_1)(e) \quad (6)$$

$$B_2(e) = (I_2 \circ L_2)(e) \quad (7)$$

$$F_{L1}(e) = (S_2 \circ L \circ S_3(B_2, L_1))(e) \quad (8)$$

In the second phase, the corresponding features are adjusted in the opposite direction and point indicators are created for operator verification:

$$F_{L2}(e) = S_3(B, L_2)(e) \quad (9)$$

$$P_L(e) = \big(P \circ M_2(B, B_2)\big)(e) \quad (10)$$

For polygon features, the first phase selects and adjusts the relevant polygon elements:

$$G(e) = (L_A \circ L_1)(e) \quad (11)$$

$$G_2(e) = (L_A \circ L_2)(e) \quad (12)$$

$$F_{G1}(e) = (S_b \circ G_2)(e) \quad (13)$$

The second phase performs boundary refinement using snapping, densification, and intersection-based correction:

$$F_{G2}(e) = Sp(D \circ Sp(L_F \circ G, PP), IP(G_2))(e) \quad (14)$$

Active Replica Only

When only the active sheet is available, the procedure prepares its boundary content for later integration with the neighbouring sheet.

For line features:

$$F_L(e) = S_4(e) \quad (15)$$

$$P_L(e) = (H \circ C)(e) \quad (16)$$

For polygon features:

$$F_G(e) = S_b(e) \quad (17)$$

Definition of Pipeline Functions

The pipeline formulas use the following variables:

– e and f represent the features that are being processed;
– ω represents an auxiliary or intermediate feature used as an additional input to an operation;
– $TK25_1$ represents the polygon of the first sheet;
– $TK25_2$ represents the polygon of the second sheet;
– b represents the distance from the sheet frame in meters;
– c represents the feature type code used for selecting specific polygon features.

The line-feature functions used in the pipeline are defined as follows:

– $I(f)$ = intersect $(f, TK25_2)$
– $I_2(f)$ = intersect $(f, TK25_1)$
– $S(f)$ = snap $(f, I_2(f))$
– $L(f)$ = select_fc_that_isnt_snaped $(f, TK25_1)$
– $S_2(f)$ = snap $(f, TK25_2)$
– $M(f, \omega)$ = merge (f, ω)
– $P(f)$ = create_changes (f)
– C = create_parallel_line_to_frame $(TK25_1, TK25_2, a)$ – where $a = b - 3$.

– $H(f)$ = create_correction_lines (f, $TK25_2$, C)
– $L_1(f)$ = select_fc_on_replica (f, $TK25_1$)
– $L_2(f)$ = select_fc_on_replica (f, $TK25_2$)
– $S_3(f, \omega)$ = snap (f, ω)
– $M_2(f, \omega)$ = merge (f, ω)
– $S_4(f)$ = snap $(f, TK25_2)$

The polygon-feature functions used in the pipeline are defined as follows:

– $L_1(f)$ = select_fc_on_replica $(f, TK25_1)$
– $L_2(f)$ = select_fc_on_replica $(f, TK25_2)$
– $L_A(f)$ = select_features (f, c)
– PP = create_parallel_line_to_frame $(TK25_1, TK25_2, b)$
– $IP(f)$ = create_intersection_fc_with_polygon (f, $TK25_1$)
– $L_F(f)$ = select_fc_on_frame_border (f, PP)
– $S_p(f,\omega)$ = snap (f, ω)
– $D(f)$ = densify_fc (f)
– $S_b(f)$ = snap $(f, TK25_1)$

The python class methods used in these functions represent the elementary geoprocessing operations from which the complete workflow is composed. In the following method definitions, *a*, *b*, and *c* are used as generic input arguments:

– intersect *(a, b)* – performs the intersection of line object *a* with polygon *b* and creates intersection points.
– snap *(a, b)* – snaps points of object *a* to standalone points, points within a line object, or boundary lines of polygon object *b*.
– select_fc_that_isnt_snapped *(a, b)* – selects line objects *a* that touch the boundary of the adjacent sheet polygon *b*.
– merge *(a, b)* – merges objects *a* and *b* into a new object.
– create_changes *(a)* – creates a point object indicating lines of object *a* that require operator verification.
– create_parallel_line_to_frame *(a, b, c)* – creates a line at distance *c*, parallel to the frame of polygon *b*, and located on sheet polygon *a*.
– create_correction_lines *(a, b, c)* – creates a point object indicating lines of object *a* that cross the auxiliary line *c*, but do not touch the boundary of the adjacent sheet polygon *b*.
– select_fc_on_replica *(a, b)* – selects line or polygon objects *a* located on sheet polygon *b*.
– select_features *(a, b)* – selects polygon objects *a* with attribute code *b*.
– create_intersection_fc_with_polygon *(a, b)* – creates intersection lines between polygon object *a* and sheet polygon *b*.
– select_fc_on_frame_border *(a, b)* – selects polygon objects *a* that intersect line *b*.
– densify_fc *(a)* – increases the number of vertices in polygon object *a*, improving the precision of subsequent snapping and boundary refinement.

This formulation presents the workflow as a structured sequence of reusable geoprocessing functions rather than as a simple execution of individual GIS tools. Such an approach makes the logic of the algorithm more transparent and provides a basis for applying the same principle to similar cartographic production tasks. This structure can also be presented through workflow diagrams, as shown in Figures 3 and 4.

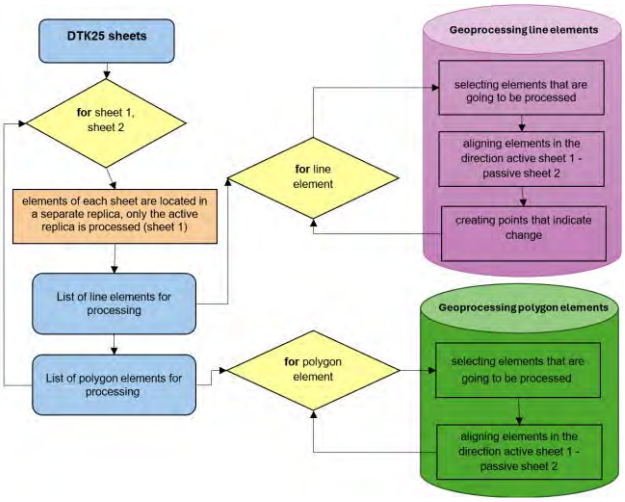


Figure 3. Diagram illustrating the content alignment workflow in the Active & Passive Replica tool

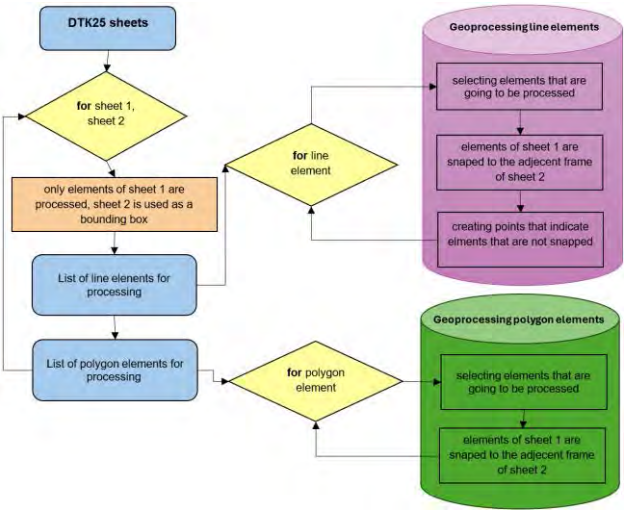


Figure 4. Diagram illustrating the content alignment workflow in the Active & Passive Replica tool

4. RESULTS AND DISCUSSION

The developed Python GIS workflow was tested on a series of adjacent DTM25 sheet pairs produced at the MGI. The main purpose of testing was to evaluate whether the developed workflow could correctly identify, refine, and control boundary geometries in different production scenarios.

The results showed that the automated processing successfully resolved most geometric inconsistencies that occurred within the defined tolerance values. In the case of line features, offsets and interruptions along sheet boundaries were corrected. The most visible improvement was observed for contour lines and road features, where manual correction is usually time-consuming because of

the number and density of elements along the map frame.

For polygon features, the workflow reduced unwanted gaps and overlaps between neighbouring boundary elements. This was especially important for large surface features that extend across sheet borders, such as vegetation areas, water surfaces, marshland, and similar polygonal elements.

The generated point control layers also proved to be useful. Instead of visually inspecting the entire sheet boundary, the operator was directed to specific locations where unresolved discrepancies or potentially problematic geometries remained. As a result, the operator's work was concentrated on complex cases that required expert assessment, while repetitive geometric corrections were performed automatically.

The effect of automation was also evaluated through the time required for sheet-boundary alignment. According to production records, the conventional manual procedure required approximately 5 working hours to align the content between two adjacent DTM25 sheets. Since one DTM25 sheet may need to be aligned with four neighbouring sheets, the total time for a complete sheet could reach approximately 20 working hours. After applying the Python GIS workflow, the average processing time was reduced to approximately 40 minutes per adjacent-sheet operation, or about 2.7 working hours for a complete sheet, including script execution, preliminary sheet analysis, and subsequent correction of marked discrepancies. This represents a time saving of more than 85% compared with the conventional procedure and an efficiency increase of more than seven times.

Overall, the results indicate that the developed Python GIS workflow provides a practical basis for semi-automated cartographic processing. The tested functions showed repeatable applicability across different feature classes and production scenarios. This confirms that the selected procedures can be performed in ArcGIS Pro while preserving the original production logic, making the workflow suitable for use in a real cartographic production environment. These results also show that the same principle can be extended to other DTM25 production tasks in which repetitive geoprocessing operations are present.

Expressing the workflow through clearly defined reusable geoprocessing functions makes the algorithm easier to analyse, test, modify, and extend. It also allows individual parts of the workflow to be adapted for other production tasks without redesigning the entire procedure.

5. CONCLUSION

Python GIS algorithms provide an effective basis for modernizing procedures in digital topographic map production. By transferring the workflow to ArcGIS Pro and Python 3, the developed solution was adapted to a current software environment while preserving the cartographic logic, production rules, and operator-guided validation procedures.

The workflow enables automatic correction of most inconsistencies in line and polygon features located along sheet borders. Generated control points allow the redirection of manual work from complete detail visual inspection of the sheet boundary to targeted verification of remaining problematic cases. In this way, automation did not remove cartographic judgement from the process. Instead, it reduced repetitive editing and focused operator attention on situations where expert control is necessary. This resulted in a substantial reduction in processing time and an efficiency increase of more than seven times.

The developed workflow establishes a methodological and practical knowledge base for applying Python programming to the automation of other processes within the production of geotopographic materials. The same logic can be adapted to other cartographic tasks, such as generalization. The results also indicate that semi-automation is an appropriate model for complex cartographic workflows. Full automation is limited by the variability of mapped content and the need for cartographic interpretation, but a carefully designed semi-automated workflow can significantly reduce manual work, improve consistency, and support a more sustainable digital topographic map production system.

References

[1] Eker, R. and Aydin, A. (2016). Customization and Automation of Geoprocessing with Python in ArcGIS for Natural Hazards and Forestry-related Applications. In Proceedings of International Forestry Symposium (IFS 2016). 07-10 December Kastamonu.

[2] Diener, M. (2015). Python Geospatial Analysis Cookbook. Birmingham: Packt Publishing Ltd. ISBN: 9781783555079

[3] Garrard, C. (2016). Geoprocessing with Python. New York: Manning Publications Co. ISBN: 9781617292149

[4] Drobnjak, S., Milonjić, Z., Đorđević, D., & Simić, M. (2025). Flash flood susceptibility mapping using LiDAR-Derived DEM and machine learning algorithms: Ljubovića case study, Serbia. Open Geosciences, 17(1). https://doi.org/10.1515/geo-2025-0859

[5] Simić, M., Vakanjac, B., & Drobnjak, S. (2025). NDGRI: A Novel Sentinel-2 Normalized Difference Gamma-Radiation Index for Pixel-Level Detection of Elevated Gamma Radiation. Remote Sensing, 17(19), 3331. https://doi.org/10.3390/rs17193331

[6] Simić, M., Vakanjac, B., Drobnjak, S., Novković, I. & Jovanović, J. (2026). Deep learning classification of elevated gamma-radiation zones using Sentinel-2 spectral data. Open Geosciences, 18(1). https://doi.org/10.1515/geo-2025-1000

[7] Zandbergen, A., P. (2020). Advanced Python Scripting for ArcGIS Pro. San Bernardino: Esri Press. ISBN: 9781589486188.

[8] Pimpler, E. (2021). Programming ArcGIS Pro with Python - 2ndEdition. Boerne: Geospatial Training Services. ISBN: 9798740349961

[9] Simić, M., Bobar, M., Bakrač, S., & Jovanović, J. M. (2025, November 20). PYTHON-BASED

AUTOMATION OF GEOPROCESSING IN DIGITAL TOPOGRAPHIC MAP PRODUCTION: CONTENT ALIGNMENT BETWEEN MAP SHEET BORDERS. 52nd Symposium on Operational Research (SYMOPIS 2025), Palić, Serbia. https://doi.org/10.5281/zenodo.17533755

[10] Dobesova, Z. (2011): Programming Language Python for Data Processing. Proceedings of International Conference on Electrical and Control Engineering (ICECE), Yichang, China, Volume 6 Institute of Electrical and Electronic Engineers (IEEE), CFP 1173J-PRT, 4866-4869 p. ISBN 978-1-4244-8163-7

[11] Toms, S. (2015). ArcPy and ArcGIS-Geospatial Analysis with Python (Vol. 1, No. 1, pp. 1-343). Birmingham, UK: Packt Publishing.

[12] Gowrishankar, S., and Veena, A. (2018). Introduction to Python Programming. Miami: CRC Press. ISBN: 9780815394372

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE EVALUATION OF SIEM PLATFORMS IN DETECTING MITRE ATT&CK TECHNIQUES BASED ON FALSE POSITIVE ANALYSIS

MARKO PREMOVIĆ
Serbian Armed Forces, Belgrade, marko.premovic@vs.rs

MLADEN VEINOVIĆ
Singidunum University, Belgrade, mveinovic@singidunum.ac.rs

BOBAN STOJANOVIĆ
Serbian Armed Forces, Belgrade, boban.stojanovic@vs.rs

IGOR PETROVIĆ
Ministry of Defence, Belgrade, igor.petrovic@mod.gov.rs

Abstract: *Security Information and Event Management (SIEM) platforms are essential for centralized security monitoring in Security Operations Centers (SOCs), yet they continue to face a persistent challenge: high volumes of false positive alerts that impair analyst efficiency and delay incident response. This paper presents a comparative evaluation of three SIEM platforms—Wazuh, IBM QRadar, and Splunk Enterprise Security—in detecting selected MITRE ATT&CK techniques. Several attack scenarios, including brute force authentication, network reconnaissance, and malicious command execution, were executed to assess detection accuracy, latency, false positive rates, and alert precision. Results reveal notable differences between open-source and commercial solutions. Enterprise platforms demonstrated superior correlation capabilities and lower false positive rates, while open-source Wazuh offered competitive detection performance at significantly reduced deployment costs. These findings provide practical insights into SIEM performance when detecting adversary behaviors mapped to the MITRE ATT&CK framework, offering guidance for organizations balancing cost, accuracy, and operational efficiency in their security stack selection.*

Keywords: *SIEM, MITRE ATT&CK, False Positives, Threat Detection, Security Monitoring*

1. INTRODUCTION

The increasing complexity of modern information systems and the continuous growth of cyber threats have significantly increased the importance of centralized security monitoring solutions. Organizations generate large volumes of security-related logs originating from endpoints, servers, authentication services, and network devices, making manual analysis of security events increasingly difficult. In such environments, Security Information and Event Management (SIEM) systems have become essential components of Security Operations Centers (SOC), enabling centralized collection, correlation, and analysis of security events in real-time [1].

Despite the widespread adoption of SIEM technologies, one of the major operational challenges remains the high number of false positive alerts generated during event correlation and threat detection processes. Excessive false positives may lead to alert fatigue, increased workload for SOC analysts, and delayed response to legitimate security incidents. Consequently, evaluating the balance between detection capability and false positive generation has become an important aspect of SIEM performance assessment.

In recent years, the MITRE ATT&CK framework has emerged as a standardized knowledge base for modeling adversarial behavior and evaluating defensive mechanisms. ATT&CK-based testing allows researchers and security analysts to simulate realistic attack techniques and objectively measure the detection capabilities of security platforms [2].

Recent SOC studies additionally emphasize that excessive alert volume may contribute to analyst fatigue, operational inefficiency, and delayed incident prioritization. Consequently, modern SIEM evaluation increasingly includes not only detection capability, but also alert quality and operational usability within SOC environments [3].

This paper presents a comparative evaluation of the Wazuh, IBM QRadar, and Splunk Enterprise Security SIEM platforms through the analysis of selected MITRE ATT&CK techniques and false positive events. The research is conducted in a controlled laboratory environment using simulated attack scenarios mapped to ATT&CK techniques, including brute force attacks, network reconnaissance, and suspicious command execution.

The primary contributions of this research are:

1) a controlled ATT&CK-based comparative evaluation of three widely used SIEM platforms under identical attack conditions;

Corresponding author: M. Premovic;
E-mail: marko.premovic@vs.rs;
DOI: 10.5937/OTEH26100P

2) quantitative assessment of false positive generation and detection effectiveness using standardized evaluation metrics;

3) operational analysis of alert quality through Precision-based evaluation of actionable alerts within SOC environments.

The paper is organized as follows. After the introduction, the second chapter presents the theoretical background on SIEM systems, the MITRE ATT&CK framework, and false positive alerts, while the third chapter describes the experimental environment, attack scenarios, and evaluation metrics. The fourth chapter presents the detection results and false positive analysis, followed by the conclusion and directions for future work.

2. THEORETICAL BACKGROUND

2.1 Security Information and Event Management Systems

Security Information and Event Management (SIEM) systems are designed to provide centralized collection, normalization, correlation, and analysis of security events generated within enterprise infrastructures.

Modern SIEM platforms utilize rule-based analytics, behavioral analysis, and threat intelligence integration to identify suspicious activities and generate alerts for SOC analysts [4].

Modern enterprise SIEM platforms additionally support integration with external threat intelligence feeds, enabling improved contextual analysis and prioritization of high-risk events during correlation processes.

However, improperly configured detection rules may generate excessive false positive events, reducing the operational efficiency of security monitoring processes.

Recent comprehensive surveys of SIEM adoption trends confirm that organizations continue to prioritize false positive reduction as a key operational challenge, with correlation rule tuning remaining one of the most resource-intensive SOC activities [13].

2.2 MITRE ATT&CK Framework

The MITRE ATT&CK framework represents a structured knowledge base used for the classification of adversarial tactics and techniques observed during cyber attacks.

One of the primary advantages of the ATT&CK framework is its applicability in detection engineering and security evaluation. By mapping attack scenarios to ATT&CK techniques, security analysts can evaluate the effectiveness of defensive mechanisms and identify potential detection gaps within monitored environments.

The ATT&CK techniques used in this research are presented in Table 1.

Table 1. Selected MITRE ATT&CK Techniques Used in the Experimental Evaluation

Technique	ATT&CK ID
Brute Force	T1110
Network Service Scanning	T1046
Command and Scripting Interpreter	T1059

2.3 False Positive Events

False positive alerts represent security events incorrectly classified as malicious activities by detection systems. High false positive rates may negatively impact SOC operations by increasing analyst workload and reducing the visibility of legitimate threats [5]. Consequently, reducing unnecessary alerts while maintaining high detection accuracy is considered one of the major challenges in SIEM optimization and detection engineering. In practical SOC environments, false positive reduction is commonly achieved through correlation rule tuning, contextual enrichment, and event prioritization mechanisms. Proper rule optimization may significantly improve analyst efficiency without reducing detection coverage [6].

3. EXPERIMENTAL ENVIRONMENT

3.1 Laboratory Setup

The laboratory environment was designed to emulate a small enterprise network consisting of approximately 50 endpoint systems. The infrastructure included Windows workstations, Linux servers, centralized authentication services, and network devices generating routine operational logs. Endpoint activity included user authentication events, administrative operations, file access events, process execution records, and network communications in order to simulate realistic enterprise workloads.

A dedicated Kali Linux system was used for attack execution, while Windows and Ubuntu hosts acted as monitored targets. Security events were collected through native operating system logging mechanisms and forwarded to the evaluated SIEM platforms. Log sources included authentication logs, process creation events, command execution records, and network connection events.

The laboratory topology was intentionally configured to generate both legitimate administrative activity and malicious behavior. This approach enabled observation of SIEM detection performance under conditions that more closely resemble operational SOC environments, where benign and malicious events coexist within the same monitoring infrastructure. Laboratory parameters are shown in Table 2.

Table 2. Laboratory Configurations Parameters

Parameter	Value
Simulated Endpoints	50
Windows Systems	Multiple
Linux Systems	Multiple
SIEM Platforms	Wazuh 4.8, QRadar 7.5, Splunk ES 7.3.2
Attack Workstation	Kali Linux
Log Sources	Auth, Process, Network, System
ATT&CK Techniques Tested	3
Repetition per Scenario	5

To ensure result comparability, all SIEM platforms received identical log streams generated from the same attack scenarios and background network activity. Platform-specific detection rules were configured according to vendor-recommended practices while maintaining comparable monitoring coverage across all evaluated solutions.

3.2 Attack Scenarios

Several attack scenarios mapped to selected MITRE ATT&CK techniques were executed during the experimental phase of this research. The scenarios were designed to represent common adversarial activities frequently encountered in enterprise environments and SOC operations. Each scenario was executed five times under identical laboratory conditions in order to evaluate detection consistency, alert generation behavior, and correlation effectiveness across the analyzed SIEM platforms.

The first scenario simulated credential access attempts using the Hydra tool and was mapped to ATT&CK technique T1110 (Brute Force). Multiple SSH authentication attempts were generated against Linux systems within the laboratory environment using valid and invalid credential combinations. The objective of this scenario was to evaluate the ability of SIEM platforms to detect repeated authentication failures and identify potential brute-force activity while minimizing unnecessary alerts generated by legitimate authentication events.

The second scenario focused on network reconnaissance activities corresponding to ATT&CK technique T1046 (Network Service Scanning). Reconnaissance was performed using the Nmap tool [8] against multiple hosts within the monitored network segment. Various scanning techniques were executed in order to generate network discovery events and service enumeration activity. This scenario was selected because reconnaissance activities often produce event patterns that resemble legitimate administrative operations, making accurate detection particularly challenging for SIEM platforms.

The third scenario simulated malicious command execution behavior associated with ATT&CK technique T1059 (Command and Scripting Interpreter). PowerShell commands commonly observed in attack investigations were executed on Windows systems in order to generate process creation events, command execution logs, and related security telemetry. The objective was to evaluate the capability of each SIEM solution to identify suspicious command execution activity while maintaining acceptable alert precision.

To provide realistic operating conditions, attack execution was performed while legitimate background activity was simultaneously generated throughout the laboratory environment. Authentication events, file access operations, system administration tasks, and normal network communications were continuously produced by the simulated endpoints. This approach enabled evaluation of SIEM performance under conditions that more closely resemble operational enterprise networks rather than isolated attack demonstrations.

Each attack scenario was repeated five times for every evaluated SIEM platform. During every execution, detection events, generated alerts, false positive occurrences, and detection latency were recorded. The collected observations were subsequently aggregated and used for the calculation of the evaluation metrics presented in this study.

3.3 Evaluation Metrics

The performance evaluation of the three SIEM platforms (Wazuh, IBM QRadar, Splunk Enterprise Security) was based on quantitative metrics derived from controlled attack simulations mapped to MITRE ATT&CK techniques T1110 (Brute Force), T1046 (Network Service Scanning), and T1059 (Command and Scripting Interpreter). Each attack scenario was repeated five times per platform, with simultaneous generation of legitimate background activity across 50 simulated endpoints to emulate realistic SOC conditions.

3.3.1 Determination of Event Classification Outcomes

For each executed attack instance and each background activity event, security events generated by the SIEM platforms were classified into four mutually exclusive categories:

True Positive (TP) -- An alert was generated, and the event corresponded to an actual executed attack technique (e.g., a Hydra brute-force SSH attempt correctly detected as T1110).

False Positive (FP) -- An alert was generated, but the event was not part of any attack; it originated exclusively from legitimate administrative or background activity (e.g., a scheduled PowerShell script flagged as suspicious command execution).

True Negative (TN) -- No alert was generated, and the event was indeed benign (e.g., a normal user login or routine file access).

False Negative (FN) -- No alert was generated, although an attack was successfully executed (e.g., an Nmap network scan that remained undetected).

Classification was performed by correlating SIEM alert logs with ground truth attack logs collected from the Kali Linux attack workstation and from target endpoints (Windows and Linux system logs). Each event was manually verified by the research team to eliminate ambiguity.

The total event set for each SIEM platform consisted of approximately 1000 benign background events (authentication, process creation, file access, network connections) and 120 attack related events (5 repetitions × 3 techniques × approximately 8 individual attack actions per technique). Thus, the total number of evaluated events per platform was 1120, distributed as TP + FP + TN + FN = 1120.

3.3.2 Detection Time Measurement

Detection Time (detection latency) was defined as the interval between the moment an attack action was completed on the target system and the moment the SIEM platform generated a corresponding alert in its console.

Detection latency was measured as the elapsed time between attack execution and alert generation within the corresponding SIEM platform. Measurements were collected for all true positive detections and averaged across all attack scenarios. The final "Avg Detection Time" reported in Table 4 represents the arithmetic mean value per platform.

3.3.3 Calculation of Accuracy and False Positive Rate (FPR)

Detection Accuracy was calculated using the standard formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Accuracy represents the proportion of all evaluated events (both malicious and benign) that were correctly classified by the SIEM.
False Positive Rate (FPR) was defined as:

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

FPR measures the ratio of benign events that were incorrectly flagged as malicious relative to all actual benign events. A lower FPR indicates better alert precision and reduced analyst fatigue.
Alert Relevance Ratio (ARR) is proposed as a complementary SOC-oriented metric:

$$ARR = \frac{TP}{TP + FP} \quad (3)$$

Precision represents the proportion of generated alerts that correspond to actual malicious activity. Higher Precision values indicate a greater proportion of actionable alerts requiring analyst investigation.

From an operational SOC perspective, Precision is additionally discussed in this paper as the Alert Relevance Ratio (ARR), emphasizing the practical value of alert quality and analyst workload reduction.

3.3.4 Raw Detection Statistics

Based on the classification methodology described in Section 3.3.1, the raw counts of TP, FP, TN, and FN were aggregated for each SIEM platform across all attack repetitions and background events. Table 3 presents these raw values.

Table 3. Raw Detection Statistics

Platform	TP	FP	TN	FN	Total Events
Wazuh	84	196	804	36	1120
QRadar	90	88	912	30	1120
Splunk	93	67	933	27	1120

Approximately 1000 benign events were generated through routine administrative activity, authentication operations, process execution, and normal network communications across the laboratory environment.

3.3.5 Derived Performance Metrics

Using the raw data from Table 3 and the equations (1) - (3), the following performance metrics were calculated for each SIEM platform.

Avg Detection Time – mean latency in seconds (measured only for TP events).

Table 4. Comparative Performance Metrics Across SIEM Platforms

Platform	Recall	(ARR)	F1	(FPR)	Accuracy	Avg Detection Time (s)
Wazuh	70%	30%	0.42	19.6%	79.3%	4.3 ± 1.2
QRadar	75%	50.6%	0.60	8.8%	89.5%	2.1 ± 0.5
Splunk	77.5%	58.1%	0.66	6.7%	91.6%	1.8 ± 0.4

Values represent the arithmetic mean across five attack repetitions. Standard deviations for detection time were as follows: Wazuh ±1.2 s, QRadar ±0.5 s, Splunk ±0.4 s, indicating consistent detection latency across attack repetitions.

The obtained metrics indicate that enterprise SIEM platforms (QRadar and Splunk) demonstrated significantly higher detection accuracy, substantially lower false positive rates, better alert relevance, and lower detection latency compared to the open-source platform Wazuh. However, Wazuh maintained acceptable detection performance with considerably reduced deployment costs, making it a viable option for organizations with limited resources.

High false positive rates were observed primarily during reconnaissance-related activities (T1046) due to the

similarity between legitimate administrative network scans and malicious scanning behavior. Such conditions may negatively affect SOC workflows by increasing alert fatigue and reducing analyst efficiency. The Alert Relevance Ratio (ARR) further highlights that in the case of Wazuh, only 30% of generated alerts were truly malicious, whereas Splunk achieved nearly 60% relevance.

4. RESULTS AND ANALYSIS

4.1 MITRE ATT&CK Detection Results

Detection performance is reported primarily using recall, precision, and their harmonic mean, since these metrics more faithfully reflect the operational reality of a SOC than accuracy under the present class distribution. The evaluation set contains 120 attack events against approximately 1000 benign events, so accuracy is dominated by the correct classification of benign traffic and overstates platform effectiveness. Splunk Enterprise Security achieves 91.6% accuracy, yet its recall is 77.5%, meaning roughly one in five attack events is missed. Recall follows the same ordering across platforms, at 70.0% for Wazuh, 75.0% for QRadar, and 77.5% for Splunk, indicating that detection coverage is broadly comparable and that the platforms separate far more sharply on alert quality than on raw coverage. Precision, reported here as the Alert Relevance Ratio, exposes that separation directly. Only 30.0% of Wazuh alerts correspond to genuine attacks, against 50.6% for QRadar and 58.1% for Splunk, so the open-source configuration imposes a substantially heavier triage burden for a similar level of coverage. The F1 score summarises both axes as 0.42, 0.60, and 0.66, respectively, confirming that Splunk dominates QRadar, which in turn dominates Wazuh, since each platform achieves both higher recall and higher precision than the platform below it. This ordering reinforces the central observation of this study, that alert relevance rather than detection coverage is the dimension on which these platforms most clearly diverge.

This finding is consistent with the base-rate fallacy problem in intrusion detection, which Axelsson [10] demonstrated nearly two decades ago, yet it remains a persistent challenge in SOC operations.

The performed experiments demonstrated that all analyzed SIEM platforms were capable of detecting the majority of executed ATT&CK techniques. However, differences were observed regarding correlation quality, alert precision, and detection latency.

Similar ATT&CK-based SIEM evaluation approaches have recently been adopted in enterprise threat detection studies [9].

In Table 5, "Partial" detection for Wazuh against T1046 (Network Service Scanning) indicates that while the platform successfully detected certain Nmap scan types (e.g., TCP SYN scan), other scan variants (e.g., UDP scan) remained undetected due to the absence of specific correlation rules in the default Wazuh configuration. In contrast, QRadar and Splunk demonstrated comprehensive detection across all scan types.

Table 5. Detection Results of MITRE ATT&CK Techniques Across Analyzed SIEM Platforms

Technique	Wazuh 4.8	QRadar 7.5	Splunk Enterprise Security 7.3.2
T1110	Detected	Detected	Detected
T1046	Partial	Detected	Detected
T1059	Detected	Detected	Detected

The obtained results indicate that enterprise platforms demonstrated improved event correlation and lower alert redundancy compared to the open-source solution [7].

4.2 False Positive Analysis

False positive analysis showed notable differences between the analyzed SIEM platforms. During network scanning scenarios, the open-source platform generated a larger number of low-priority alerts, while enterprise platforms applied more precise correlation mechanisms that reduced unnecessary event generation.

The Precision-Recall curves in Figure 1 illustrate the trade-off between detection precision and recall across different classification thresholds. Splunk Enterprise Security consistently achieves higher precision than QRadar for the same recall level, while Wazuh exhibits the lowest precision across the entire recall range. This confirms that commercial platforms provide better alert quality, which directly reduces the number of false positives requiring analyst review.

High false positive rates were observed primarily during reconnaissance-related activities due to the similarity between legitimate administrative operations and suspicious scanning behavior. Such conditions may negatively affect SOC workflows by increasing alert fatigue and reducing analyst efficiency.

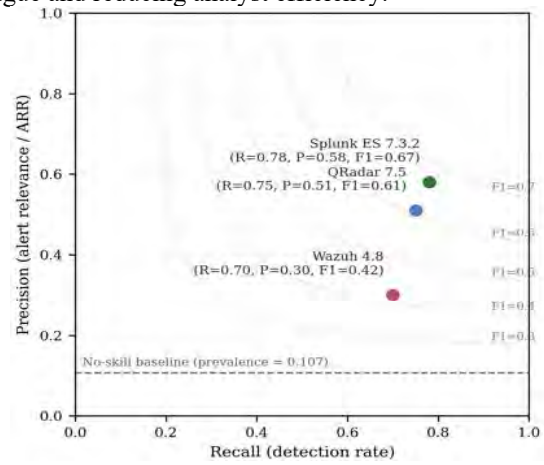


Figure 1. Precision-Recall Curves for the Evaluated SIEM Platforms

The operational impact of false positive rates can be illustrated through a typical SOC scenario. Assuming a daily volume of 1,000 benign events requiring analysis, Wazuh's FPR of 19.6% would generate approximately

196 false positive alerts per day, whereas Splunk's FPR of 6.7% would generate only 67 false positives. This difference of 129 additional alerts per day contributes directly to analyst fatigue and may delay the investigation of true positive incidents. Consequently, the choice of SIEM platform has measurable implications for SOC staffing requirements and mean time to respond (MTTR). The operational impact of false positive alerts on SOC workflows has been extensively documented in recent studies [11].

Beyond the numerical false positive rate, an important observation was the operational impact of alert clustering. Some reconnaissance-related false positives appeared as isolated events, while others triggered multiple correlated alerts for the same activity. Although this did not directly affect detection accuracy, it increased the number of events requiring analyst review. This observation aligns with recent qualitative research by Alahmadi et al. [12], who found that SOC analysts perceive the majority of security alerts as non-actionable, contributing to alert fatigue and operational inefficiency.

4.3 Performance Evaluation

Performance analysis included resource utilization and average detection latency during attack execution. The enterprise platforms demonstrated improved scalability and faster event correlation under higher event volumes, while the open-source platform required lower deployment costs and hardware resources.

The comparative analysis indicates that the selection of a SIEM platform strongly depends on organizational requirements, available infrastructure, and operational priorities.

It should be noted that the presented results were obtained within a controlled laboratory environment with a limited number of attack scenarios and endpoints. Real-world enterprise infrastructures may introduce additional variables, including encrypted traffic, heterogeneous logging formats, and significantly larger event volumes.

Another limitation of the presented evaluation is the relatively small number of ATT&CK techniques included in the testing process. While the selected techniques represent commonly observed adversarial behaviors, modern attack campaigns often combine multiple tactics across different stages of the intrusion lifecycle. Consequently, the obtained results primarily reflect platform performance under focused attack scenarios rather than full attack-chain simulations. Future evaluations may benefit from incorporating multi-stage attack emulation in order to assess SIEM correlation capabilities under more complex operational conditions.

5. CONCLUSION

This paper presents a comparative evaluation of the Wazuh, IBM QRadar, and Splunk Enterprise Security SIEM platforms through the analysis of selected MITRE ATT&CK techniques and false positive events. The conducted experiments demonstrated that all analyzed platforms were capable of detecting common adversarial activities, including brute force attacks, network reconnaissance, and suspicious command execution.

The obtained results showed that enterprise SIEM solutions provided improved event correlation and reduced false positive rates, while the open-source platform offered competitive detection capabilities with significantly lower deployment costs. Furthermore, the research highlighted the operational impact of false positive alerts on SOC workflows [9] and emphasized the importance of correlation rule optimization in modern security monitoring environments [3][6].

ARR is proposed as a complementary indicator for future SIEM benchmarking studies.

The study also indicates that alert relevance and correlation quality may represent equally important evaluation criteria as traditional detection metrics when assessing SIEM effectiveness in operational SOC environments.

Based on the obtained results, organizations with limited budgets and smaller infrastructures may benefit from Wazuh's competitive detection performance at reduced costs, while large enterprises requiring high precision and low latency should consider commercial solutions such as Splunk Enterprise Security or IBM QRadar.

Future research may include larger-scale infrastructures, additional ATT&CK techniques, and integration of machine learning approaches for adaptive SIEM correlation and false positive reduction.

The findings suggest that alert quality metrics may provide greater operational insight than detection accuracy alone. This is consistent with the base-rate fallacy problem identified by Axelsson [10], who demonstrated that accuracy becomes misleading when attacks constitute a small fraction of total events, as is typical in enterprise environments.

References

[1] K. Kent, M. Souppaya, "Guide to Computer Security Log Management," NIST Special Publication 800-92, National Institute of Standards and Technology, 2006. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf DOI: 10.6028/NIST.SP.800-92

[2] MITRE ATT&CK Team, “MITRE ATT&CK: Design and Philosophy,” MITRE Corporation, 2020. Available: https://attack.mitre.org/

[3] Splunk Threat Research Team, “The State of Security Operations: Alert Fatigue and SOC Efficiency,” Splunk Research Report, 2024. Available: https://www.splunk.com/

[4] S. Peisert, M. Bishop, "Improving Security Event Correlation in Modern SOC Environments," IEEE Security & Privacy, vol. 20, no. 3, pp. 44-52, 2022. DOI: 10.1109/MSEC.2022.3154325

[5] A. Herrero, E. Corchado, "False Positive Reduction in Intrusion Detection Systems," International

Symposium on Distributed Computing and Artificial Intelligence,pp. 253-258, 2009.
DOI: 10.1007/978-3-642-02481-8_34

[6] Elastic Security Labs, “Detection Engineering and SIEM Rule Tuning Practices,” Elastic Security Research, 2023. Available: https://www.elastic.co/security-labs/

[7] G. Lyon, “Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning,” Insecure.Com LLC, 2009.

[8] A. Ahmed, K. Kim, "Evaluation of SIEM Platforms for Enterprise Threat Detection Using MITRE ATT&CK," Computers & Security, vol. 128, 2023. DOI: 10.1016/j.cose.2023.103185

[9] M. C. D'Amico, J. Whitley, "The Real Cost of Alert Fatigue in Security Operations Centers," IEEE Security & Privacy, 2023. DOI: 10.1109/MSEC.2023.3289472

[10] S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," *ACM Transactions on Information and System Security*, vol. 3, no. 3, pp. 186-205, 2000, doi: 10.1145/357830.357849.

[11] B. A. Alahmadi, L. Axon, and I. Martinovic, "99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms," in Proc. 31st USENIX Security Symposium, 2022, pp. 2783-2800.

[12] G. González-Granadillo, S. González-Zarzosa, and R. Díaz, "Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures," *Sensors*, vol. 21, no. 14, art. 4759, 2021, doi: 10.3390/s21144759.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MODULARITY OR COMPRESSION? A UNIFIED COMPARISON OF MIXTURE-OF-EXPERTS AND POLICY DISTILLATION AS EXECUTION MODULES FOR LLM-PLANNED AGENTS

IGOR CIGANOVIĆ
School of Computing, Union University, Belgrade, iciganovic217draf.rs

ALEKSANDAR PLUŠKOSKI
School of Computing, Union University, Belgrade, apluskoski317d@raf.rs

MILOŠ JOVANOVIĆ
Technical Test Center, Belgrade, milosjovanovic1967@gmail.com

JELENA VASILJEVIĆ
School of Computing, Union University, Belgrade, jvasiljevic@raf.rs

Abstract: *Hierarchical agents that delegate long-horizon planning to a large language model (LLM) and low-level control to a learned execution module have recently been shown to solve sparse-reward tasks that defeat monolithic reinforcement learning (RL) agents. Two distinct designs of the execution module have been proposed independently for the same environment: a mixture-of-experts (MoE) architecture that retains one specialized policy per sub-task and routes between them, and a single network obtained by distilling several expert policies into one. Both were reported to outperform an end-to-end baseline, but they have never been compared to each other, and the published numbers were obtained under different expert-training procedures, which makes a direct reading of them unsafe. This paper places both designs in a single architectural framework, re-analyses the published results under a common notation, and identifies precisely which comparisons the existing evidence does and does not support. We find that the two execution modules are separated by 0.02 average reward, far below the resolution of single-seed experiments, while the expert policies they were built from differ by up to 18 percentage points on identical sub-tasks purely because of how they were initialized. We further show that under both designs the planner contributes essentially no errors, so the execution module is the binding constraint in both systems.*

Keywords: *reinforcement learning, hierarchical agents, mixture of experts, policy distillation, large language models, MiniGrid, sparse reward*

1. INTRODUCTION

Reward sparsity remains one of the practical limits on applying reinforcement learning (RL) to tasks of realistic length [1], [2]. When the environment returns a scalar reward only after a long sequence of interdependent actions has been executed correctly, the credit-assignment problem becomes intractable for a single monolithic policy, and training either stagnates or converges to a policy that solves only the shortest missions. The difficulty compounds with horizon: the probability of a randomly explored trajectory completing every required sub-task falls off geometrically with the number of sub-tasks.

A family of recent architectures attacks this problem by refusing to learn planning and control in the same network. Planning, which requires abstract reasoning over a long horizon, is delegated to a large language model (LLM) that is queried once per episode and returns an ordered list of sub-goals. Control, which requires short-horizon navigation and manipulation, is left to an execution module trained by standard RL on individual sub-tasks, each of which carries a sufficiently dense reward to be learnable. Because the LLM never participates in training and is not queried at every step, the training pipeline remains a conventional RL procedure, and the inference cost of the language model is amortized over an entire episode.

Within this shared skeleton, two different designs of the execution module have been proposed independently and evaluated in the same extended MiniGrid environment. The first keeps the specialists separate: one PPO policy is trained per sub-task, and a gating network performs hard routing, selecting exactly one expert whose output is forwarded as the system action [3]. The second compresses the specialists: the trained experts act as teachers, generate a dataset of state-policy pairs, and a single student network is trained by supervised policy distillation to reproduce their behaviour across all sub-tasks [4]. One design proposes that capacity should stay partitioned, and the other that it can be shared without loss.

Both were reported to outperform an end-to-end PPO agent on multi-step missions, and both used the same planner, the same environment generator, the same reward

Corresponding author: I. Ciganović;
E-mail: iciganovic217d@raf.rs
DOI: 10.5937/OTEH26101C

function, and the same PPO implementation. Neither study set out to compare the two against each other, and neither had reason to: each was answering whether its own design beat a flat baseline. Having built both, we are now in a position to ask the question that neither could. The two designs imply opposite engineering decisions about memory footprint, inference cost, and the effort required to extend the agent with a new skill, and a practitioner choosing between them currently has no evidence to consult.

This paper does not report new training runs. Its contribution is to make the existing evidence legible and to establish what it can and cannot support. Concretely, we (i) express both architectures in one notation and identify the exact point at which they diverge, (ii) re-tabulate all published results under a common metric and flag the confounds that prevent a direct reading, (iii) show that the two execution modules are separated by an effect far smaller than the experimental resolution that produced them, (iv) demonstrate that in both systems the planner is not the limiting component.

2. RELATED WORK

2.1. Sparse reward and hierarchical control

Deep reinforcement learning handles dense-reward control well [5]; the difficulty here is specific to sparse reward. Classical remedies for reward sparsity include curriculum learning, which orders sub-tasks from easy to hard while retaining a single policy [6], and temporal abstraction, which introduces explicitly represented sub-policies. The options framework [7] and its differentiable successor, the option-critic architecture [8], learn both the sub-policies and the termination conditions end to end. FeUdal Networks [9] and h-DQN [10] instead separate a manager that emits goals from a worker that reaches them. These methods share the ambition of the present work, but they learn the decomposition jointly with the policy, which reintroduces a difficult optimization problem precisely where the reward is sparsest.

2.2. Mixture of experts

The mixture-of-experts formulation dates to Jacobs et al. [11], where a gating network combines the outputs of specialized sub-networks. Modern large-scale variants use sparse gating to activate a small subset of experts per input, yielding conditional computation at constant cost [12], [13]. The architecture considered here occupies the extreme of that spectrum: gating is hard, exactly one expert is active, and the experts are complete policies trained in isolation rather than sub-networks of a jointly optimized model.

2.3. Knowledge and policy distillation

Distillation transfers the behaviour of one or more teachers into a student by supervised regression on the teachers' outputs [14]. In the RL setting, policy distillation [15] and actor-mimic [16] showed that several task-specific agents can be compressed into a single network, often with improved stability, because the student learns from a fixed target distribution rather than from non-stationary environment interaction. The variant used in the system studied here is multi-teacher distillation with an important additional degree of freedom: the student may be initialized either randomly or from a policy already trained by RL on all sub-tasks simultaneously.

2.4. Comparing modular and monolithic policies

Direct comparisons between modular and compressed policies are rare, partly because the two are usually proposed as answers to different questions: mixtures of experts for capacity scaling, distillation for deployment efficiency. Where both appear in one study, the comparison is typically incidental. The RL literature also lacks agreed conventions for equalizing the training budget of a modular system against a monolithic one, since the natural unit of accounting, environment steps, is consumed per policy rather than per system. This paper treats that accounting question as a first-class obstacle rather than a detail.

2.5. Language models as planners

Following the emergence of instruction-following language models [17], [18], several systems have used them to decompose tasks for embodied agents. Zero-shot planners extract action sequences directly from an LLM [19]; SayCan grounds the language output in learned affordances [20]; ReAct interleaves reasoning and acting [21]; Voyager maintains a growing skill library in an open-ended environment [22]. Systems that query the model at every step incur an inference cost proportional to episode length, and systems that place the model inside the training loop inherit that cost multiplied by the number of training episodes. The architecture examined here queries the model exactly once per episode, before the first action.

3. UNIFIED ARCHITECTURE

We describe both systems as instances of a single three-stage architecture, which allows the point of divergence to be stated precisely. Figure 1 shows the shared pipeline and the two execution variants.

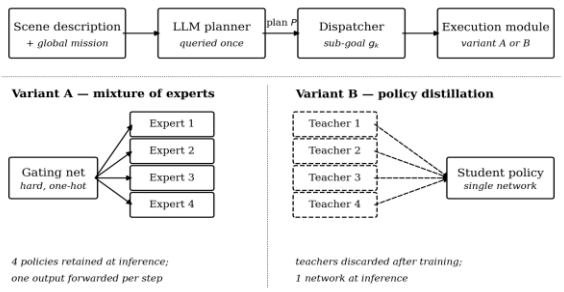


Figure 1. The unified three-stage architecture. Planner and dispatcher are shared; the two systems differ only in the execution module, which either routes between retained specialists (variant A) or replaces them with a single distilled student (variant B).

3.1. Formal setting

The environment is a partially observable Markov decision process. At step t, the agent receives an observation consisting of a 7x7x3 egocentric view, a scalar direction, and a tokenized mission string, and emits an action a_t from the discrete MiniGrid action set. A mission M is solvable by an ordered plan P = (g_1, ..., g_K) of sub-goals drawn from a fixed vocabulary G = {go to goal, go to <object>, toggle <object>, pick up <object>}. The reward is granted only on completion of the global mission and is discounted by the number of steps taken,

$$R = 1 - 0.9\frac{num_steps}{max_steps} \quad (1)$$

so that a perfect score of 1.0 is unattainable and all reported values are bounded well below it.

3.2. Stage 1: planner

A structured textual description of the scene, generated by the environment during scene construction, is combined with the global mission and submitted once to a locally hosted LLM. The model returns an ordered list of sub-goals from G. The planner is not queried again for the remainder of the episode. It takes no part in training, and its parameters are frozen.

3.3. Stage 2: dispatch

Sub-goals are issued to the execution module one at a time as the current mission string. Control returns to the dispatcher when the execution module emits the done action, at which point the next sub-goal in P becomes active. No external oracle verifies that the sub-goal was in fact achieved; the module's own done signal is the only completion event. If the plan is exhausted or the step budget expires, the episode terminates.

3.4. Stage 3: execution, variant A (mixture of experts)

Variant A maintains four independent policies, one per sub-goal type in G, each trained by PPO in isolation on episodes whose entire mission is that sub-goal. A gating network, consisting of an embedding layer, a GRU, and a linear layer, consumes the current mission string and emits a one-hot vector selecting the active expert. Formally, the module computes

$$f(x) = sum_i(w_i(x)f_i(x)) \quad (2)$$

With w_i representing the weight corresponding to the i-th expert, and f_i the output of the said expert. w(x) is constrained to the vertices of the simplex, so exactly one expert's output survives. The gating network is trained by supervised learning on the fixed vocabulary.

3.5. Stage 3: execution, variant B (policy distillation)

Variant B trains the same four experts, then discards them at inference time. Each expert is rolled out in the environment to produce a dataset of pairs (o, p_i_expert(o)), the datasets are pooled, and a single student network with the architecture of one expert is trained by supervised learning to reproduce the pooled action distributions. Two initializations are considered: a randomly initialized student, and a student initialized from a PPO policy previously trained on all sub-tasks simultaneously, referred to below as the all model. Parameter count is constant in the number of skills, and no routing decision is made at inference time.

3.6. Cost model

The two variants differ in a way that is visible before any experiment is run. Let θ denote the parameter count of one expert policy, γ that of the gating network, and N the number of skills. Variant A stores N*θ + γ parameters and must keep all N policies resident, since the identity of the next sub-goal is not known until the dispatcher issues it. Variant B stores θ regardless of N. With the network described in Section 4.2 and N = 4, variant A therefore carries roughly four times the execution-time footprint of variant B, plus a small gating network dominated by its 32 x 1024 embedding layer.

The per-step compute tells a different story. Because gating is hard, exactly one expert is evaluated per step, so the forward-pass cost of variant A equals that of variant B plus one gating evaluation, which is amortized because the gate need only be recomputed when the sub-goal changes. The asymmetry is therefore in memory, not in arithmetic. Extending the agent with a new skill costs one additional PPO training run and one additional resident policy under variant A, against one additional teacher plus a full re-distillation of the student under variant B. Neither published study measured any of these quantities, and Section 5.5 records them as open.

3.7. Baselines

The end-to-end model, denoted no LLM, is a single PPO policy trained directly on complete multi-step missions with no planner at all; it isolates the benefit of the hierarchy itself.

4. EXPERIMENTAL SETUP

Because this paper re-analyses previously published experiments, the setup below is the one common to both source studies [3], [4]. Where the two studies diverge, the divergence is stated explicitly, since it is the substance of Section 5.4.

4.1. Environment

MiniGrid v3.0.0 [23] was extended to allow control over mission complexity. The grid is fixed at 11x11, giving at most four rooms. Object placement is constrained so that generated scenes are always solvable and never obstruct the agent, the goal tile, or a doorway. Doors carry colors and may be locked, in which case a key of the same color is required; keys may be placed inside boxes, adding a level of nesting. The environment additionally emits a

structured textual description of the scene for the planner. Training and evaluation scenes were sampled from the same generator but drawn as disjoint sets.

4.2. Policy and training

All policies use the same feature extractor: a three-branch network processing the image, direction, and mission inputs, whose latent vectors are concatenated and fed to actor and critic heads. Training uses PPO [2] as implemented in Stable-Baselines3 [24], with hyperparameters obtained by population-based bandit search [25]. Each run consists of 20 million steps with 16 parallel environments, and training proceeds in consecutive runs with decreasing learning rate and epoch count until improvement stops. The number of runs was chosen per policy by that stopping rule, not fixed in advance.

4.3. Planner configuration

The planner is qwen3:30b served locally through Ollama at temperature 0.0 with a fixed system prompt describing the available instructions, objects, colors, and the key-door rule. We benchmarked alternative open models on a fixed set of scenes; the selected model matched all reference solutions, while gemma2:27b, deepseek-r1:32b, and llama3:8b matched 13, 11, and 9, respectively [4].

4.4. Metrics

Two quantities are reported. Average reward is the mean of R over an evaluation set of 1000 freshly sampled episodes. Success rate is the fraction of episodes in which the global mission was completed, computed on a manually verified subset of 100 episodes and disaggregated by the number of sub-tasks in the plan. Sub-task-level results are reported as the average reward obtained by a policy when the sub-task is the entire mission.

5. RESULTS

This section re-tabulates the published results of both systems under the common notation of Section 3.

5.1. Sub-task level

Table 1 reports the performance of individual policies on each sub-task. Two facts stand out. First, specialization is clearly beneficial: every specialist outperforms the shared all model on its own sub-task, and every specialist scores exactly zero on sub-tasks it was not trained for, confirming that the experts carry no incidental cross-task competence. Second, and more consequentially, the pick-up and toggle experts differ substantially between the two studies. In [3] they reach 0.57 and 0.47; in [4], where they were fine-tuned from the all model rather than trained from scratch, they reach 0.68 and 0.65.

Table 1. Sub-Task Performance Of Individual Policies (Average Reward)

Policy	Goal	Object	Pick	Togg
Go to goal	0.86	0.00	0.00	0.00
Go to object	0.00	0.72	0.00	0.00
Pick up [3]	0.00	0.00	0.57	0.00
Toggle [3]	0.00	0.00	0.00	0.47
Pick up [4]*	0.00	0.00	0.68	0.00
Toggle [4]*	0.00	0.00	0.00	0.65
All (shared)	0.75	0.65	0.59	0.58

* fine-tuned from the all model rather than trained from scratch.

The gap matters. On two of four sub-tasks, the distillation system started from teachers 11 to 18 points stronger than the experts the MoE system routes between. Neither paper had cause to flag this, since neither was comparing across the two, but any end-to-end comparison that ignores it inherits it as a confound.

5.2. Multi-task execution

Table 2 reports execution modules evaluated across all sub-tasks simultaneously, without a planner. The ordering is informative. Distillation from a randomly initialized student reaches 0.56, which is worse than the all model it was meant to improve upon; distillation from the all model reaches 0.67, which is better. The identity of the initialization, rather than the distillation procedure itself, accounts for a swing of 0.11.

The MoE module has no directly comparable entry in this table, because hard routing makes its aggregate multi-task performance a weighted average of the per-expert values of Table 1 rather than a separately measured quantity. Computing that average over the four sub-task types gives an implied value of approximately 0.66, but this is an arithmetic construction and not a measurement, and it is reported here only to indicate the order of magnitude.

Table 2. Execution Modules Evaluated Across All Sub-Tasks, No Planner

Execution module	Reward	Seeds
All (PPO, shared policy)	0.65	1
Distilled, random init	0.56	1
Distilled, from all model	0.67	1
MoE, hard routing	~0.66*	1

* implied weighted average of Table 1, not an independent measurement.

5.3. End-to-end performance

Table 3 reports complete systems on multi-step missions.

Both hierarchical systems improve substantially on both baselines: against the end-to-end no-LLM agent the margin is 0.12 to 0.14 average reward, and against a non-hierarchical PPO agent it is 0.22 to 0.24. The benefit of delegating planning is therefore not in question.

The comparison between the two execution modules, however, is. The MoE system reports 0.57 and the distilled system 0.55, a difference of 0.02. All figures come from single runs at a fixed seed, which was a reasonable choice when the comparison of interest was against a flat baseline separated by 0.12 or more. It is not enough resolution for a 0.02 gap. RL training at fixed hyperparameters varies across seeds, and [3] reports exactly such a spread in final expert performance over four seeds. We read Table 3 as saying the two execution modules are indistinguishable on the evidence available, rather than as a narrow win for MoE.

Table 3. End-To-End Performance On Multi-Step Missions (1000 Episodes)

System	Reward	Seeds
PPO	0.33	1
End-to-end, no LLM	0.43	1
LLM + distilled [4]	0.55	1
LLM + MoE [3]	0.57	1

5.4. Complexity and the location of failure

Table 4 disaggregates success rate by plan length for the MoE system and its baselines. Performance degrades sharply with horizon: from 98.44 percent on single-sub-task missions to 28.57 percent at three sub-tasks and 14.29 percent at four. The end-to-end agent, by contrast, collapses to near zero as soon as more than one sub-task is required, solving one three-sub-task episode only through an incidental behaviour in which an object lying on the path happened to be the required key.

Crucially, manual verification of 100 planner outputs found no invalid plans in either study; every plan was a correct, if not always optimal, solution to the global mission. Since the planner introduced no failures, all observed degradation is attributable to the execution module. This holds for both variants and is the single most important shared finding: the two systems differ in how they organize execution, and execution is exactly what limits them.

A rough consistency check supports this. Treating sub-task failures as independent, the expected success rate of a three-step pick up, toggle, go-to-goal sequence under the MoE expert rates of Table 1 is 0.57 x 0.47 x 0.86, or about 0.23, against 0.2857 observed. The estimate is close enough to make execution failure the plausible dominant mechanism, though the independence assumption is untested and the observed value exceeds the prediction, which suggests failures are in fact mildly correlated with scene difficulty rather than independent.

Table 4. Success Rate By Plan Length (Percent) calculated on the first 100 episodes

System	1 sub	3 sub	4 sub
PPO	68.75%	0.00%	0.00%
End-to-end, no LLM	68.75%	4.76%	7.14%
LLM + distilled [4]	89.06%	23.81%	14.29%
LLM + MoE [3]	98.44%	28.57%	14.29%
sample size	64	21	14

5.5. Planner accuracy and cost

Because the planner is shared, its properties bound both systems equally. Table 5 reports the model-selection benchmark from [4]. The selected model matched every reference solution on the benchmark set, and two of the three alternatives were close behind, which indicates that on 11x11 scenes the planning task is near saturation for models of this size. That is convenient for the present comparison, since it holds the planner constant, but it also means the benchmark cannot discriminate between planners and gives no evidence about behaviour on larger scenes.

Planning cost, by contrast, is substantial. Reported inference time is 3 to 15 minutes per episode on a single consumer GPU, dominated by the thinking phase of the model, against a per-step execution cost measured in milliseconds. The hierarchy therefore trades a large fixed cost at the start of each episode for a much cheaper trajectory, and any deployment argument that turns on latency must account for this term, which is identical for both execution modules and larger than either.

Table 5. Planner Model Selection Benchmark [4]

Model	Matched	Params
qwen3:30b (selected)	14 / 14	30B
gemma2:27b	13 / 14	27B
deepseek-r1:32b	11 / 14	32B
llama3:8b	9 / 14	8B

Benchmark set of 14 reference scenes. Near-saturation at this scene complexity limits the discriminative power of the benchmark.

5.6. What the evidence does not support

Table 6 summarizes the comparisons required to choose between the two execution modules, together with their current status. Of seven, three are settled by published data, and four are open. None of the published runs report either module at a matched step budget, at matched parameter count, over multiple seeds, or with latency and memory figures. That is not an oversight so much as a consequence of scope: those quantities only become interesting once the two designs are placed side by side, which is what we are

doing here for the first time.

Table 6. Status Of The Comparisons Required To Choose A Module

Comparison	Status
Beats end-to-end baseline	Settled, both
Planner error contribution	Settled, both
Success by plan length	Settled, both
Matched step budget	Open
Matched parameter count	Open
Multi-seed with CIs	Open
Inference latency, memory	Open

6. DISCUSSION

Three things come out of this.

The hierarchy is settled. The execution module is not. Delegating planning to a language model queried once per episode produces a large and consistent improvement over end-to-end RL on multi-step missions, reproduced independently in both studies with different execution modules. We take that as settled. The choice between routing specialists and compressing them rests on a single 0.02 difference measured once, and we would treat it as open.

Initialization matters more than the distillation procedure does. The gap between distilling from random weights and distilling from a trained policy is 0.11, five times the gap between the two architectures. A randomly initialized student ends up worse than the model that would have supplied its initialization. That points to imitation of expert action distributions not being enough on its own to build the representation, and suggests distillation works here as a consolidation step on an already competent policy rather than as a way of training one from nothing.

Execution is what binds in both designs. Across the two studies, the planner produced no errors in 100 manually verified episodes, so every route to a better system runs through the execution module. That changes the question worth asking. Not which module scores higher today, but which improves faster per unit of extra training, and which degrades more gracefully once the planner starts making mistakes. A planner with no observed errors on 11x11 grids will not stay error-free on larger scenes.

The two designs are not just different implementations of the same capability. They fail differently, and that is testable. A routed mixture degrades sub-task by sub-task, since a weak expert damages only the plans that invoke it, and the remaining experts are untouched. A distilled student shares parameters across all skills, so degradation is expected to be diffuse, and a poorly represented sub-task may be compensated or overwritten by the others. The published data are consistent with this reading.

The published evidence supports a weak recommendation, and it follows from the architecture rather than from the reward numbers. Where the skill set is fixed and small, memory is not constrained, and skills are added rarely, MoE architecture is preferable because a new skill costs one training run and cannot disturb existing ones. Where the agent must run on constrained hardware, or where the skill set is large enough, a distilled network is preferable, provided the student is initialized from a multi-task policy rather than randomly. The 0.11 penalty for random initialization is the one effect in this analysis large enough to act on with confidence.

Resolving the comparison needs nothing conceptually new. Both modules should be trained under a declared step-accounting rule, reported under both conventions, since a four-expert MoE granted the same per-expert budget as a single student consumes four times the total experience. Both should be evaluated over at least five seeds with confidence intervals, on at least 200 episodes per complexity bracket, and reported alongside parameter count, memory footprint, and per-step inference latency.

7. REPRODUCIBILITY

All numbers re-analysed in this paper originate from the two source studies [3], [4], both of which published their source code and a video demonstration of the resulting agents. The environment is a modified MiniGrid v3.0.0, the RL implementation is Stable-Baselines3 v2.7.0, and the planner is served through Ollama v0.30.0 with the configuration reproduced in Section 4.3. Training was performed on a single consumer GPU, with each 20-million-step run requiring between four and seven hours of wall-clock time.

We note one obstacle to exact reproduction that affects both studies equally. The number of consecutive training runs per policy was determined by an improvement-based stopping rule applied during training rather than fixed in advance, so the total number of environment steps consumed by each published policy is a function of the observed learning curve. Reproducing the published numbers therefore requires reproducing that decision process, not merely the hyperparameters.

8. CONCLUSION

We have placed two independently proposed execution modules for LLM-planned hierarchical agents, a hard-routed mixture of experts and a multi-teacher distilled policy, into a single architectural framework and re-analysed their published results under common notation. The hierarchical design itself is well supported: both variants improve on an end-to-end PPO agent by 0.12 to 0.14 average reward on multi-step missions, and in both the planner was found to contribute no failures, making the execution module the binding constraint.

The comparison between the modules, however, is not supported. They are separated by 0.02 average reward from single-seed runs, while the experts underlying them differ by up to 18 percentage points on identical sub-tasks

solely because of initialization, and the effect of initialization on the distilled student is itself five times larger than the architectural difference. We conclude that the choice between modularity and compression remains open.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction, 2nd ed*, MIT Press, 2018.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "*Proximal policy optimization algorithms,*" arXiv, 2017.

[3] I. Ciganović, A. Pluškoski, M. Jovanović, and J. Vasiljević, "*Evaluation of multi-model architecture against single-model PPO in the MiniGrid environment,*" Electronics, vol. 15, no. 14, art. 3170, 2026.

[4] A. Pluškoski, I. Ciganović, M. Jovanović, and J. Vasiljević, "*Benefits of using LLM for long-term planning with distilled subtask model compared to end-to-end reinforcement learning in the MiniGrid simulator,*" Electronics, vol. 15, no. 9, art. 1921, 2026.

[5] V. Mnih et al., "*Human-level control through deep reinforcement learning*", Nature, vol. 518, no. 7540, pp. 529–533, 2015.

[6] S. Narvekar et al., "*Curriculum learning for reinforcement learning domains: A framework and survey,*" arXiv, 2020.

[7] R. S. Sutton, D. Precup, and S. Singh, "*Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning,*" Artificial Intelligence, vol. 112, no. 1–2, pp. 181–211, 1999.

[8] P.-L. Bacon, J. Harb, and D. Precup, "*The option-critic architecture,*" in Proc. AAAI, 2017, pp. 1726–1734.

[9] A. S. Vezhnevets et al., "*FeUdal networks for hierarchical reinforcement learning,*" in Proc. ICML, 2017, pp. 3540–3549.

[10] T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum, "*Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation,*" in Proc. NeurIPS, 2016.

[11] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "*Adaptive mixtures of local experts,*" Neural Computation, vol. 3, no. 1, pp. 79–87, 1991.

[12] N. Shazeer et al., "*Outrageously large neural networks: The sparsely-gated mixture-of-experts layer,*" in Proc. ICLR, 2017.

[13] W. Fedus, B. Zoph, and N. Shazeer, "*Switch Transformers: Scaling to trillion parameter models with simple and efficient sparsity,*" J. Mach. Learn. Res., vol. 23, no. 120, pp. 1–39, 2022.

[14] G. Hinton, O. Vinyals, and J. Dean, "*Distilling the knowledge in a neural network,*" in NIPS Deep Learning Workshop, 2015.

[15] A. A. Rusu et al., "*Policy distillation,*" in Proc. ICLR, 2016.

[16] E. Parisotto, J. L. Ba, and R. Salakhutdinov, "*Actor-mimic: Deep multitask and transfer reinforcement learning,*" in Proc. ICLR, 2016.

[17] A. Vaswani et al., "*Attention is all you need,*" in Proc. NeurIPS, 2017.

[18] T. Brown et al., "*Language models are few-shot learners,*" in Proc. NeurIPS, 2020.

[19] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch, "*Language models as zero-shot planners: Extracting actionable knowledge for embodied agents,*" in Proc. ICML, 2022.

[20] M. Ahn et al., "*Do as I can, not as I say: Grounding language in robotic affordances,*" in Proc. CoRL, 2022.

[21] S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," in Proc. ICLR, 2023.

[22] G. Wang et al., "*Voyager: An open-ended embodied agent with large language models,*" Trans. Mach. Learn. Res., 2024.

[23] M. Chevalier-Boisvert et al., "*Minigrid & Miniworld: Modular and customizable reinforcement learning environments for goal-oriented tasks,*" in Proc. NeurIPS Datasets and Benchmarks, 2023.

[24] A. Raffin et al., "*Stable-Baselines3: Reliable reinforcement learning implementations,*" J. Mach. Learn. Res., vol. 22, no. 268, pp. 1–8, 2021.

[25] J. Parker-Holder, V. Nguyen, and S. J. Roberts, "*Provably efficient online hyperparameter optimization with population-based bandits,*" in Proc. NeurIPS, 2020.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

AI-ASSISTED DECISION SUPPORT IN COMMAND-AND-CONTROL SYSTEMS: APPLYING THE SENSE–UNDERSTAND–DECIDE–ACT (SUDA) CYCLE

IVAN VULIĆ, MILJAN VUČETIĆ, ILIJA POPADIĆ, RANKO PETROVIĆ, ZORAN BANJAC
Vlatacom Institute of High Technologies, Milutina Milankovića 5, Belgrade, Serbia

Abstract: *Modern command-and-control (C2) systems face a paradox: the volume of sensor data grows exponentially while the time available for decision-making shrinks. This paper presents an architectural framework for AI-assisted decision support in C2 systems based on the Sense–Understand–Decide–Act (SUDA) cycle and demonstrates it on a deployed system: the Vlatacom C7ISTAR solution, in which the ORAO command platform carries the ARES (Algorithmic Reconnaissance and Execution System) three-tier AI core, in operational service protecting national critical infrastructure since 2024. Each phase is mapped to a technology layer and characterized by measured quantities: a panoramic thermal sensor covering the full circle in 2.3 s; an edge-AI stage above 100 TOPS with sub-100 ms inference latency, reducing about fifty raw detections per second to approximately one plausible target; and an agile electro-optical turret classifying a cued object in under two seconds. We argue that decision advantage arises not from replacing the human decision-maker but from restructuring the C2 information flow around the SUDA cycle, so that the path from sensor to decision-maker, and from decision to effect, becomes short, traceable and measurable.*

Keywords: *command and control, decision support, artificial intelligence, SUDA cycle, counter-UAS.*

1. INTRODUCTION

Sensors have proliferated across every domain of the modern battlefield—radar, electro-optical and infrared imaging, radio-frequency and acoustic sensing, unmanned platforms, and networked reporting—so a contemporary command post receives orders of magnitude more data than any human staff can assimilate in the time available. At the same time the velocity of the threat has increased: unmanned systems in large numbers, loitering munitions, electronic attack and cyber effects compress the interval between first detection and required response from hours to minutes, and in terminal defense to seconds.

Hence the central paradox of this paper: commanders have never had more information, and never less time in which to use it. The side that resolves it obtains decision advantage — the ability to close its own decision cycle faster, and with better judgment, than the adversary. The classical articulation is Boyd's observe–orient–decide–act (OODA) loop [1], refined for C2 analysis by Brehmer [2]; Endsley's theory of situation awareness [3] captures the quality dimension. Command-support terminology has evolved in step: from C2 through C4ISR — within which geographic information systems first brought spatial decision support to the operational picture [4] — to the extended C5ISR, C6ISTAR and C7ISTAR designations that successively admit cyber defense, combat systems, and collaboration and interoperability [5]; C7ISTAR, the fullest of these, is adopted here as the designation of the unified sovereign command-and-mission portfolio that the described solution implements.

Artificial intelligence can now materially contribute to both the speed and the quality of the cycle, yet its integration into C2 raises questions as much doctrinal as technical — how much authority may be delegated to a machine, how a recommendation is explained to the operator who must act on it, how models operating on classified data are kept under sovereign control — and recent scholarship agrees that acceleration is valuable only while human judgment remains authoritative within the loop [6]. Beneath these lies the practical question this paper takes as its organizing principle: how long is the path from sensor to decision-maker, what happens to the information along the way, and how much of the journey can be entrusted to a machine.

The paper defines the SUDA cycle and positions it against the OODA loop, situation-awareness theory and the JDL data-fusion model (Section 2); maps each phase onto a technology layer (Section 3); instantiates the framework on the deployed Vlatacom C7ISTAR solution, with measured quantities per phase (Section 4); analyses human–machine teaming and the constraints of edge inference and sovereignty (Sections 5 and 6); and traces a complete engagement with its cycle-time budget (Section 7). Section 8 concludes.

2. THE SUDA CYCLE AS A C2 FRAMEWORK

The SUDA cycle structures a C2 system's information flow into four continuously repeating phases. In Sense, data are acquired from all available sources. In Understand, they are fused into a single coherent representation — the common operational picture (COP) — and enriched with machine-derived interpretation: classification, anomaly detection, and short-term prediction. In Decide, the system generates feasible courses of action (COA), evaluates them against constraints, and presents them to the human decision-

Corresponding author: I. Vulić;
E-mail: ivan.vulic@vlatacom.com
DOI: 10.5937/OTEH26102V

maker. In Act, the selected course is translated into tasking of effectors and sensors, execution is monitored, and observed effects feed the next iteration; Figure 1 summarises the cycle.

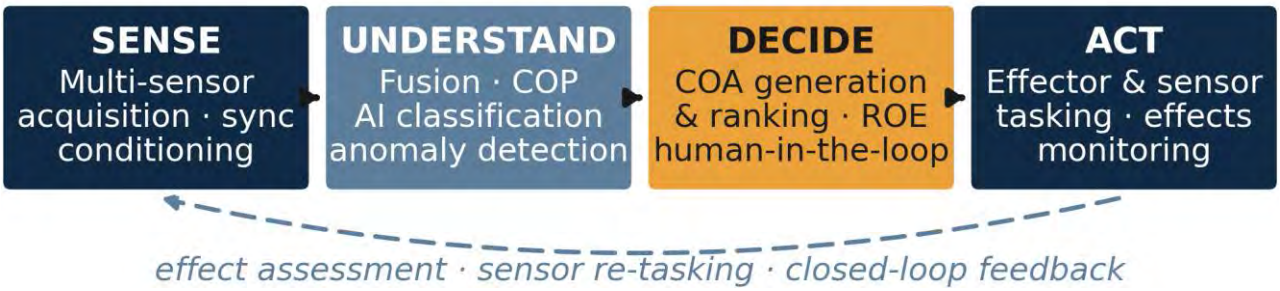


Figure 1. The SUDA decision cycle with the technology layer associated with each phase

SUDA deliberately restates Boyd's loop in system-engineering form: Observe maps to Sense; Orient — the phase Boyd considered decisive — to Understand, where fusion and machine learning are concentrated; Decide and Act retain their names. The gain is mapping discipline: each phase becomes an identifiable set of services with interfaces and latency budgets, and cycle time — from a change in the real world to a completed response — becomes a key performance indicator, decomposable into per-phase terms that Section 4 measures on a deployed system.

The framework aligns with established theory and doctrine. Endsley's three levels of situation awareness [3] span Sense and Understand; the JDL data-fusion model [7] structures the Understand phase—historically the bottleneck of the C2 cycle and therefore where AI offers the greatest compression —the central design thesis of this paper. The Joint All-Domain Command and Control (JADC2) concept likewise organizes capability around sense, make sense, and act — matching Sense, Understand, and the combined Decide–Act — and requires machine analysis to be explainable, confidence-tagged, and accessible only to authorized users [8], [5]. SUDA is thus not a competitor to these formulations but their engineering expression.

3. MAPPING SUDA PHASES TO TECHNOLOGY LAYERS

Figure 2 presents the functional architecture resulting from mapping the SUDA phases onto technology layers: read upwards from the base, it is the sensor-to-decision-maker path drawn as a stack, with the cryptographic rail protecting the observation at every station.

3.1. Sense — multi-sensor acquisition

The Sense layer comprises the heterogeneous sensor set — radar, electro-optical and infrared imaging, radio-frequency and acoustic sensors, unattended ground sensors — together with structured reporting from deployed units and, increasingly, open-source (OSINT) feeds. Three engineering requirements dominate. First, time synchronization and geo-registration: fusion is only as good as the spatial and temporal alignment of its inputs, so every observation must carry a trusted timestamp and position reference. Second, standardized metadata: sensor identity, mode, confidence, and provenance must accompany every detection so the evidence can be weighed correctly; that trust begins with authenticating the sensors themselves, for which the authors' earlier Internet-of-Battlefield-Things model provides a basis [9]. Third, data conditioning at the edge, so that constrained links carry information rather than raw redundancy. The layer must also degrade gracefully: losing individual sensors or links should reduce coverage, not collapse the picture.

3.2. Understand — fusion, the COP and AI services

The Understand layer transforms conditioned observations into shared meaning. Its foundation is multi-sensor fusion in the sense of the JDL model [7]: associating detections into tracks, correlating tracks across sensors, and maintaining a single consistent COP that all echelons consume. Fusion may occur at the detection level or the track level; practical systems combine both, trading communication load against fusion gain [10], and at the semantic end of this range the Institute has applied fuzzy-relational methods to fuse heterogeneous surveillance data [11]. The integrity of the COP's geospatial foundation must be protected in its own right [12].

On top of the fused picture, AI services operate: classification models assign object types; anomaly-detection models learn an area's pattern of life and flag deviations that escape threshold-based rules; predictive models project track behavior, supporting projection in Endsley's sense [3]. Two properties are essential: every machine-derived statement carries explicit uncertainty — confidence and provenance, surfaced to the user — and the services are modular and retrainable, components with a lifecycle.

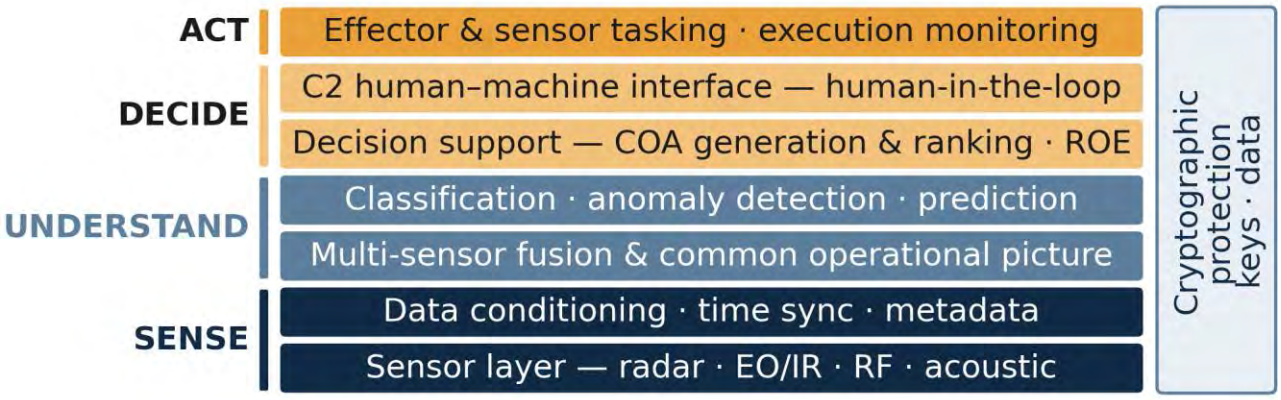


Figure 2. Functional architecture of AI-assisted C2, organized by SUDA phase, with cryptographic protection as a cross-cutting layer

3.3. Decide — decision support with the human in the loop

The Decide layer converts understanding into options: threat evaluation, COA generation, and ranking against explicit criteria — effectiveness, resource consumption, risk, deconfliction and, critically, compliance with the rules of engagement (ROE), encoded as machine-readable constraints so that no proposed option violates them by construction; what-if simulation lets the operator probe an option before committing. Large language models have recently been demonstrated as generators of candidate courses of action at machine speed [13]; the same literature documents their unreliability under adversarial and escalatory conditions [14] — which is precisely why this layer is designed so that the machine proposes and the human disposes. Ranked options are presented with their rationale — factors, contributing sensors, confidence — and every recommendation and human selection is logged, creating the audit trail on which accountability and institutional learning depend.

3.4. Act — tasking, monitoring and feedback

The Act layer translates the selected course of action into tasking of effectors and, equally, of sensors: closing the loop includes directing collection where confirmation of effect is needed. Execution is monitored against the plan, and effect assessment — whether the action achieved the intended result — is fed back into Sense and Understand, updating the picture and, over time, the models themselves. In a realistic system, many SUDA cycles run concurrently, from seconds in point defense to hours at the operational level.

4. REALISATION: THE ORAO C7ISTAR PLATFORM

The framework of Section 3 is realized in ORAO, the command-and-control platform of the Vlatacom Institute (the name is proper to this software platform, unrelated to the aircraft): a scalable micro-services system that integrates remote sensing equipment, maintains a common operational picture and lets the operator act in the field remotely. ORAO has been in operational service since 2024, protecting a national critical energy-infrastructure installation against small commercial and FPV drones — the class current counter-UAS surveys identify as hardest to counter at such sites [15], [16]. It rests on a decade of the Institute's research across the SUDA layers: electro-optical platforms and video processing [17], [18], object tracking and AI-based detection [19], [20], [21], fusion of heterogeneous sensor data [11], and cryptographic protection of operational data [22], [23].

ARES, the Algorithmic Reconnaissance and Execution System, supplies the platform's intelligence— a modular AI core conceived as a software brain over the existing sensor base, organized in three tiers that map almost one-to-one onto the SUDA phases. ARES Edge infers on a processing module at the sensor itself, classifying targets in under a second without transmitting a high-bandwidth video stream — the transition from Sense to early Understand, viable on constrained or intermittent networks. ARES Core is the operational tier: it fuses encrypted telemetry of the order of ten thousand events per second from several sites into a single COP and applies a unified ontology that recognizes behavioral patterns rather than isolated events, carrying Understand and preparing Decide. ARES Strategic is the command tier, coupling the picture to large-language-model agents through which a commander obtains reports and candidate courses of action by voice or text — the pattern demonstrated by [13], here on an operational picture rather than a simulation, under the human authority and audit trail of Section 3.3. Across all tiers, AI is combined with rule-based events and actions in a supervised autonomous mode — responses deployed automatically or manually per configuration and countermeasure type — and, being microservice-based, the platform adds sensors, effectors, and functions without altering the core, spanning fixed and vehicle-mounted configurations alike. Figure 3 shows the platform fusing radio-frequency detection with electro-optical tracking of the same target while displaying sensor and effector coverage on the map.

Each SUDA phase carries measured figures, summarised in Table 1 — engineering values read from the system's own instrumentation and logs during commissioning and routine operation, quoted to characterize the architecture, not as contractual guarantees. The Sense layer illustrates the data-volume problem concretely: a thermal panoramic sensor of 1280 × 1024 pixels, driven by a high-end video-

processing platform developed at the Institute [18], covers the full circle in 36 sectors every 2.3 s. The Understand layer illustrates the compression AI achieves: the raw flow of about fifty candidate detections per second, dominated by clouds and birds, is reduced to approximately one plausible target per second by tracking, cloud and bird filtering, and direct referral of close targets to AI classification — techniques building on the Institute's robust tracking and learned detection in thermal and short-wave-infrared imagery [19], [20], [21] — a fifty-to-one reduction before any human sees the picture: the difference between an operator who is overwhelmed and one who is informed.

Figure 3. The ORAO platform in operation at a demonstration site: the live optical stream with telemetry, the map with bearing lines from RF spectrum sensing, and the zone and sector overlay of sensors and effectors on a single console (identifying details redacted)

Table 1. SUDA phases instantiated on the ORAO platform, with characteristic measured quantities

SUDA phase	Function	Realisation	Characteristic figures
Sense	Multi-sensor acquisition, conditioning, time reference	Thermal panoramic head; RF spectrum sensing; vMSIS3 EO/IR turret with laser rangefinder; a dedicated passive acoustic sensor; X-band radar where permitted	Full circle in 2.3 s; RF spectrum 20 MHz–6 GHz
Understand	Track fusion, COP, AI classification	ARES Edge and ARES Core on current-generation GPU; classes: drone, bird, aircraft, helicopter; ontology layer; Kalman filtering with deep-learning re-identification	Above 100 TOPS; inference latency below 100 ms; four parallel video streams; 50 raw detections/s reduced to about 1/s
Decide	Threat evaluation, option preparation, ROE constraints	ORAO decision services with ARES Strategic: AI and rule-based events and actions; LLM-generated COA; supervised autonomous operation	Operator authority by default; automatic response only where explicitly configured
Act	Effector and sensor tasking, monitoring, feedback	Cueing of the vMSIS3 turret; RF jamming with selective radiation and coverage control; effect assessment returned to the COP	Turret classifies a cued object in under 2 s from any previous position

The Decide and Act layers are equally concrete. Rules of engagement appear as configuration: the protected installation applies selective control of jammer radiation power and coverage zones so that countermeasures do not reach the surrounding populated area — an ROE constraint compiled into system behavior exactly as Section 3.3 requires — and safeguards prevent countermeasures from triggering without authorization unless a fully automatic mode has been explicitly enabled. The Act layer closes the loop within the electro-optical turret's two-second verification window, automatically directed to coordinates supplied by another sensor and returning a classified, ranged, and geo-located object to the picture that cued it; the Institute has reported this turret's optical platform and pointing performance [17].

5. HUMAN–MACHINE TEAMING

Whether AI-assisted C2 delivers value depends less on model accuracy than on the quality of the human–machine team; three requirements structure the problem. The first is a deliberate allocation of function. The taxonomy of Parasuraman, Sheridan and Wickens [24] distinguishes automation of information acquisition, analysis, decision selection and action implementation, each at graded levels. In the SUDA framing, high automation is appropriate in Sense and much of Understand, where machines outperform humans in throughput and vigilance — the fifty-to-one reduction of Section 4 is no human task at all. In Decide, automation prepares and ranks options, but authority over the use of force remains with the human commander; delegation beyond that point must be explicit, bounded, and reversible—the standard allied policy now codifies autonomy in weapon systems [25]. Autonomy is not fixed: control tightens when time permits, and narrowly defined responses are pre-authorized where reaction times make deliberation impossible — supervised autonomous operation in ORAO is precisely a graduated setting of this parameter.

The second is explainability. A recommendation that cannot be explained will in practice be either ignored or blindly followed — both failure modes. The field systematized in the DARPA XAI program [26] offers

many techniques, but the operational requirement is specific: explanations in the operator's terms — which evidence, from which sensors, drove this ranking; what the confidence is; what would change the conclusion — delivered at the moment of decision and honest about uncertainty, the transparency recent work identifies as the foundation of operator situation awareness in human-AI teams [27]. A classification shown on the console with its confidence value alongside the imagery behind it is a modest but real instance of the principle.

The third is trust calibration: over-trust produces automation bias — operators accepting machine output uncritically, a failure mode documented across military decision-support deployments [28] — while under-trust wastes the machine's contribution. Calibration is built through transparency of confidence, training in the system's characteristic strengths and failure modes, and consistent logging. Responsibility is not transferred to software: the commander remains accountable, and the system must be designed so that this accountability is real, traceable, and fair.

6. IMPLEMENTATION CONSIDERATIONS

6.1. Edge and central inference

A practical architecture distributes intelligence between the sensor edge and the center: latency-critical inference (detection, first-line classification, data conditioning) at the edge, where it also relieves communication links; resource-intensive functions (model training, large-scale fusion, long-horizon analytics) centrally. In the deployed system, all AI processing occurs locally, in the field or vehicle, so the system remains operational without a network connection, and the edge platform is retrained locally on newly collected data. The edge–center connection assumes contested communications: edge nodes continue autonomously on locally cached models when links fail, the COP degrades gracefully, and synchronization resumes when connectivity returns. Adoption in an existing C2 system is likewise incremental: the layer first enters as an advisory overlay through standardized interfaces — including NATO STANAG and legacy national formats — and services migrate to integrated status as confidence and accreditation mature, building the doctrinal and trust foundations of Section 5.

6.2. Security and sovereignty

AI-assisted C2 operates on data whose confidentiality, integrity and availability are matters of national security. Data in transit and at rest are therefore protected by cryptographic means under national control — centralized key management, an interoperable public-key infrastructure and secure key distribution [29]; the chain's trustworthiness rests on key generation, hence hardware true-random generation as the foundation — the Institute's research on autonomous symmetric ciphering and key distillation [22], [23] addresses exactly this chain. The C2 network also defends itself with the same family of methods it offers the commander: machine-learning detection of anomalies in the network's own behavior [30]. Models are assets: provenance known, training data controlled, updating performed within the national security perimeter — models trained on classified operational data cannot depend on external infrastructure — the deeper reason why local retraining at the edge is an architectural requirement, not merely a latency optimization. Emission security is addressed through hardware design and accreditation. Sovereign control over data, models and cryptography is thus a precondition of operational trust.

7. CYCLE-TIME ANALYSIS OF AN ENGAGEMENT

Consider a protected critical-infrastructure site of the kind described in Section 4 — roughly one square kilometer, observed by a thermal panoramic head, RF spectrum sensing and an agile electro-optical turret, with a countermeasure tower under selective radiation control, integrated through ORAO; the objective is to stop small drones before they approach within 500 m of the perimeter. In Sense, the panoramic head sweeps the full circle every 2.3 s while the RF layer listens across the spectrum; a drone entering the outer zone is registered by one or both — an emitter by its control link, a silent one by its thermal signature against the sky. In Understand, the raw flow is compressed before an operator sees anything: detections are bound into tracks, cloud and bird returns rejected, and the surviving track fused with the RF bearing into one object in the COP, carrying position, classification, confidence, and contributing sensors. ORAO cues the turret; its edge-AI returns a class — drone rather than bird, aircraft or helicopter — with a confidence value in under two seconds, while the laser rangefinder converts bearing and range into coordinates: a few seconds after first detection, the operator has one identified, located, ranked object instead of fifty ambiguous stimuli per second.

In Decide, the platform prepares the response: which countermeasure applies, which zone it may radiate into, whether the rules permit automatic action or require approval—and, because the ROE are compiled into the configuration, it offers no option that would radiate into the prohibited sector. The operator, alerted by a single track rather than a continuous stream, reviews the classification and supporting imagery, then authorizes the measure. In Act, the jammer radiates its selective pattern, the turret observes, and the outcome — the drone descending, returning or landing — is recorded as effect assessment, refining both the picture and the models. The same discipline scales beyond soft-kill: with layered gun-based and directed-energy effectors from the Institute's portfolio, the chain from first radar contact with a multi-drone raid at 8 km to confirmed neutralization closes within about 45 seconds.

The interval from first stimulus to a directed, compliant response is thus compressed from the timescale of manual correlation across separate consoles to that of a supported decision, with every step transparent, logged and under human authority; and it decomposes — 2.3 s of panoramic revisit in Sense; sub-100 ms inference and under 2 s of turret verification in Understand; a human decision bounded by how well the option was prepared; a tasking latency in Act — each term a separate engineering problem with a separate owner. That is what it means to say that

SUDA turns cycle time from a metaphor into a measurable system property.

8. CONCLUSION

This paper has presented an architectural framework for AI-assisted decision support in C2 systems, organized around the SUDA cycle and demonstrated on a system in operational service. Its central claim is that decision advantage arises not from replacing the human decision-maker but from restructuring the C2 information flow so that machines do what they do best — acquire, fuse, classify, detect anomalies, prepare and rank options — while human judgement is applied, deliberately and accountably, where it matters most. The Vlatacom C7ISTAR realisation, in which the ORAO platform carries the ARES three-tier AI core, substantiates the claim phase by phase with the measured figures of Table 1. Together those figures describe a single object — a short, instrumented and accountable path from sensor to decision-maker and from decision to effect — and it is that path, not any individual algorithm, that modern command support has to engineer. The sovereign, end-to-end integration of sensors, intelligence and effectors within one platform is becoming the defining requirement of the field.

Future work follows three lines: standard metrics and instrumentation for per-phase cycle-time measurement across heterogeneous systems; test, evaluation and accreditation of learning-based functions, including behaviour under distribution shift and adversarial conditions; and interoperability of SUDA-structured decision support across systems and coalition boundaries.

Acknowledgement

The research underlying the sensing and AI components of this paper was funded in part by the Vlatacom Institute of High Technologies (project no. 157, VMSIS3_Advanced) and by the Science Fund of the Republic of Serbia (Grant no. 6524745, AI-DECIDE).

References

[1] BOYD,J.R.: A Discourse on Winning and Losing (ed. Hammond,G.T.), Air University Press, Maxwell AFB, 2018.

[2] BREHMER,B.: The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control, Proc. 10th ICCRTS, McLean, VA, 2005.

[3] ENDSLEY,M.R.: Toward a Theory of Situation Awareness in Dynamic Systems, Human Factors, 37(1) (1995) 32–64.

[4] VULIĆ,I.: Mobilni geografski informacioni sistemi u sistemima C4I2 (Mobile Geographical Information Systems in C4I2 Systems), Vojnotehnički glasnik, 54(3) (2006) 354–367.

[5] VULIĆ,I.: Komandno-informacioni sistemi (Command-Information Systems), University of Defence in Belgrade, Belgrade, 2026 (in Serbian, in press).

[6] JOHNSON,J.: Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command-and-Control Decision-Making in the Digital Age, Defence Studies, 23(1) (2023) 43–67.

[7] STEINBERG,A.N., BOWMAN,C.L., WHITE,F.E.: Revisions to the JDL Data Fusion Model, Proc. SPIE 3719, 1999, pp.430–441.

[8] U.S. DEPARTMENT OF DEFENSE: Summary of the Joint All-Domain Command and Control (JADC2) Strategy, Washington, D.C., 2022.

[9] VULIĆ,I., PRODANOVIĆ,R., TOT,I., BOGIĆEVIĆ,D.: Model for Authenticating the Internet of Military Things and Internet of Battlefield, Proc. ICIST 2020, Kopaonik, 2020, pp.206–210.

[10] HALL,D.L., LLINAS,J.: An Introduction to Multisensor Data Fusion, Proceedings of the IEEE, 85(1) (1997) 6–23.

[11] MOSTAFA,M.A.R.M., VUČETIĆ,M., STOJKOVIĆ,N., LEKIĆ,N., MAKAROV,A.: Fuzzy Functional Dependencies as a Method of Choice for Fusion of AIS and OTHR Data, in: Smart Sensors and Devices in Artificial Intelligence, Springer, Basel, 2022.

[12] VULIĆ,I., BORISOV,M., PRODANOVIĆ,R., RANČIĆ,D., PETROVIĆ,V.M., STANKOVSKI,S., OSTOJIĆ,G.: Protection of Digital Elevation Model — One Approach, Applied Sciences, 12(19) (2022) 9898.

[13] GOECKS,V.G., WAYTOWICH,N.: COA-GPT: Generative Pre-trained Transformers for Accelerated Course of Action Development in Military Operations, arXiv:2402.01786, 2024.

[14] RIVERA,J.-P., MUKOBI,G., REUEL,A., LAMPARTH,M., SMITH,C., SCHNEIDER,J.: Escalation Risks from Language Models in Military and Diplomatic Decision-Making, Proc. ACM FAccT '24, Rio de Janeiro, 2024.

[15] PARK,S., KIM,H.T., LEE,S., JOO,H., KIM,H.: Survey on Anti-Drone Systems: Components, Designs, and Challenges, IEEE Access, 9 (2021) 42635–42659.

[16] SEMENYUK,V., KURMASHEV,I., LUPIDI,A., et al.: Advances in UAV Detection: Integrating Multi-Sensor Systems and AI for Enhanced Accuracy and Efficiency, Int. J. Critical Infrastructure Protection, 49 (2025) 100744.

[17] LATINOVIĆ,N., POPADIĆ,I., TOMIĆ,B., SIMIĆ,A., MILANOVIĆ,P., NIJEMČEVIĆ,S., PERIĆ,M., VEINOVIĆ,M.: Signal Processing Platform for Long-Range Multi-Spectral Electro-Optical Systems, Sensors, 22(3) (2022) 1294.

[18] POPADIĆ,I., PERIĆ,M., TOMIĆ,B., SIMIĆ,A., MILANOVIĆ,P.: High-End Video Processing Platform for Electro-Optical Systems, Proc. OTEH 2020, Belgrade, 2020.

[19] PAVLOVIĆ,M., BANJAC,Z., KOVAČEVIĆ,B.: Object Tracking in SWIR Imaging Based on Both Correlation and Robust Kalman Filters, IEEE Access, 11 (2023) 63834–63851.

[20] LATINOVIĆ,N., VUKOVIĆ,T., PETROVIĆ,R., PAVLOVIĆ,M., KADIJEVIĆ,M., POPADIĆ,I., VEINOVIĆ,M.: Implementation Challenge and

Analysis of Thermal Image Degradation on R-CNN Face Detection, Telfor Journal, 12(2) (2020) 98–103.

[21] PAVLOVIĆ,M., MILANOVIĆ,P., STANKOVIĆ,M., PERIĆ,D., POPADIĆ,I., PERIĆ,M.: Deep Learning Based SWIR Object Detection in Long-Range Surveillance Systems: An Automated Cross-Spectral Approach, Sensors, 22(7) (2022) 2562.

[22] MILOSAVLJEVIĆ,M., RADOMIROVIĆ,J., UNKAŠEVIĆ,T., BOŽILOVIĆ,B.: One Class of Ideally Secret Autonomous Symmetric Ciphering Systems Based on Wiretap Polar Codes, Mathematics, 12(23) (2024) 3724.

[23] RADOMIROVIĆ,J., MILOSAVLJEVIĆ,M., BANJAC,Z., JOVANOVIĆ,M.: Secret Key Distillation with Speech Input and Deep Neural Network-Controlled Privacy Amplification, Mathematics, 11(6) (2023) 1524.

[24] PARASURAMAN,R., SHERIDAN,T.B., WICKENS,C.D.: A Model for Types and Levels of Human Interaction with Automation, IEEE Trans. SMC — Part A, 30(3) (2000) 286–297.

[25] U.S. DEPARTMENT OF DEFENSE: DoD Directive 3000.09, Autonomy in Weapon Systems, Washington, D.C., 2023.

[26] GUNNING,D., AHA,D.W.: DARPA's Explainable Artificial Intelligence (XAI) Program, AI Magazine, 40(2) (2019) 44–58.

[27] ENDSLEY,M.R.: Supporting Human-AI Teams: Transparency, Explainability, and Situation Awareness, Computers in Human Behavior, 140 (2023) 107574.

[28] PROBASCO,E., TONER,H., BURTELL,M., RUDNER,T.G.J.: AI for Military Decision-Making: Harnessing the Advantages and Avoiding the Risks, CSET, Georgetown University, 2025.

[29] PRODANOVIĆ,R., VULIĆ,I.: Model za PKI interoperabilnost u Republici Srbiji (A Model for PKI Interoperability in Serbia), Vojnotehnički glasnik, 65(2) (2017) 530–549.

[30] PROTIĆ,D., STANKOVIĆ,M., PRODANOVIĆ,R., VULIĆ,I., et al.: Numerical Feature Selection and Hyperbolic Tangent Feature Scaling in Machine Learning-Based Detection of Anomalies in the Computer Network Behavior, Electronics, 12(19) (2023) 4158.

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION VII

Materials and technologies - MT

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SOLITON-MEDIATED ENERGY TRANSFER AND MAGNETIZATION CONTROL IN PLD-GROWN E-FE_2O_3 THIN FILMS

VIOLETA N. NIKOLIĆ
Vinča Institute of Nuclear Sciences „Vinča", University of Belgrade, Mike Petrovića Alasa 12-14, P.O. Box 522, 11001 Belgrade, violeta@vin.bg.ac.rs

JOSE F. M. L. MARIANO
Universidade do Algarve, FCT, Campus de Gambelas, 8005-139 Faro, Portugal
Center of Physics and Engineering of Advanced Materials (CeFEMA), IST, Universidade de Lisboa, Av. Rovisco Pais, 1096-001 Lisboa, Portugal, jmariano@ualgt.pt

Abstract: *Laser-induced synthesis of ε-Fe_2O_3 thin films via pulsed laser deposition (PLD) involves complex energy transfer processes, impacted by coupled exciton–spin solitons in the precursor α-Fe_2O_3 lattice. This paper is continuation of our previous work, presenting a combined analytical and numerical investigation of soliton decoupling thresholds and the resulting magnetization in ε-Fe_2O_3. A soliton-based model describes energy transfer from laser photons to exciton and spin–spin solitons, allowing derivation of threshold photon energies for decoupling, as a function of laser fluence, exposure time, and exciton–spin coupling. Numerical simulations using a MATLAB-based solver produced two-dimensional and three-dimensional plots, revealing nonlinear and saturating dependencies of soliton stability on laser parameters. Furthermore, a Magnetization Calculator was developed to quantify the normalized magnetization, incorporating additional factors, such as: damping, pulse duration, and soliton amplitude. The combination of analytical thresholds and numerical modeling provides a framework for exploring PLD synthesis conditions and their potential influence on the magnetic response of ε-Fe_2O_3 thin films.*

Keywords: *ε-Fe_2O_3, exciton–spin soliton, soliton decoupling, pulsed laser deposition, magnetization modeling*

1. INTRODUCTION

ε-Fe_2O_3 thin films are of interest because of their strong magnetic anisotropy, high coercivity, and high-frequency ferromagnetic resonance, making them promising for high-frequency magnetic and electromagnetic applications [1–5]. However, the metastable ε-phase is difficult to stabilize during synthesis, and its magnetic properties depend sensitively on growth conditions [6–8]. Pulsed laser deposition (PLD) provides a route for controlling these conditions, but the laser-induced energy-transfer processes underlying the formation and magnetic response of ε-Fe_2O_3 remain insufficiently understood.

Various approaches have been proposed to improve ε-Fe_2O_3 thin film synthesis, particularly using pulsed laser deposition (PLD) [9–11]. In our previous study, theoretical optimization of PLD synthesis parameters was performed, to improve the properties of ε-Fe_2O_3 thin films. Ref. [12] investigated the impact of laser choice on the magnetization of the final product. Using soliton theory, the study analyzed the coupling between exciton solitons and spin-spin solitons, induced by laser irradiation of the α-Fe_2O_3 target during PLD. This paper is continuation of the study, presented in Ref. [12], providing deeper insight in the laser-induced energy-transfer pathway, for which is assumed that impact the magnetic ordering and stability of the final product phase of the PLD synthesis. Performed numerical computations and MATLAB simulations illustrate the applicability of the model and provide a software tool to assist experimental design in PLD synthesis, facilitating the production of ε-Fe_2O_3 thin films with improved magnetization and functional properties.

Consequently, it has potential 2.THEORETICAL MODEL

In Ref. [12], a theoretical model based on soliton theory [13], was proposed to describe energy transfer from the laser-excited α-Fe_2O_3 precursor to the ε-Fe_2O_3 product. According to this model, exciton solitons are formed through interactions between laser photons and the electron subsystem, while spin-spin solitons arise due to spin interactions in the ferromagnetic chains of the hematite crystal lattice.

Solitons are localized energy packets capable of carrying information through the lattice. Consequently, a coupled pair of exciton and spin-spin solitons can transfer energy from the precursor to the final ε-Fe_2O_3 lattice.

To gain deeper insight into the amount of laser photon energy, which is required for decoupling of the coupled pair of exciton soliton and spin-spin soliton, and building on the exciton soliton amplitude provided in Ref. [12], in this study is introduced the corresponding exciton soliton energy. Based on the cubic-quintic nonlinear Schrödinger equation (CQNLSE) considered in Ref. [12], the standard soliton energy can be expressed as the integral of the Hamiltonian density over the spatial coordinate z:

Corresponding author: V. Nikolić;
E-mail: violeta@vin.bg.ac.rs
DOI: 10.5937/OTEH26103N

$$E_{\text{exciton}} = \int_{-\infty}^{\infty} \left[M_L \left(\frac{\partial \rho}{\partial z}\right)^2 - \frac{G_L}{2}\rho^4 - \frac{F_L}{2}\rho^6 \right] \mathrm{dz} \quad (1)$$

where is E_{exciton}=$\mathrm{M}_L \left(\frac{\partial \rho}{\partial z}\right)^2$ – gradient energy (analogue to the kinetic-like term), and $\frac{-G_L}{2}\rho^4 - \frac{F_L}{2}\rho^6$ – cubic and quintic nonlinearities (analogue to the potential-like term) [14].

This integral can be solved analytically, resulting in an expression relating soliton energy to amplitude, showing that higher amplitudes correspond to higher energies—reflecting stronger energy transfer under higher laser excitation [14, 15], eq. (2):

$$E_{\text{exciton}} \sim \mathrm{M}_L \frac{\rho_0^2}{L} - \frac{G_L}{2}\rho_0^4 L - \frac{F_L}{3}\rho_0^6 L \quad (2)$$

where is soliton width given by expression: L=$\frac{4M_L}{G_L \rho_{0^2}(1+\mathrm{B}_L)}$.

Considering eq. (10) from Ref. [12], and eq. (2), the energy of the coupled pair of exciton soliton and spin-spin soliton is given by eq. (3):

$$\tilde{H}_{coupled} = \tilde{E}_{exc} + \tilde{E}_{spin} + \lambda \tilde{E}_{exc} + \tilde{E}_{spin} \quad (3)$$

where is: $\tilde{H} = \frac{H}{E_0}$, and $\tilde{E} = \frac{E}{E_0}$. To maintain dimensional consistency, the energies in eq. (3) are expressed relative to a characteristic energy scale, E0, in order to perform dimensional normalization. The exciton-spin-soliton coupling coefficient (describing the nonlinear coupling between the excitonic and spin-soliton contributions), λ, is consequently dimensionless, and describes the strength of the nonlinear energetic interaction between the exciton-soliton and spin-soliton contributions. The coupled Hamiltonian illustrates the total energy of the exciton soliton and spin-spin soliton system, including their mutual energy transfer.

Efficient magnetic energy transfer requires that the two soliton subsystems remain coupled during transfer from precursor to product. If the exciton soliton energy exceeds a characteristic coupling energy, decoupling occurs, limiting magnetization transfer. This defines a threshold condition separating regimes of effective and ineffective energy transfer.

Let us recall that the exciton soliton energy is described by eq. (4):

$$E_{\text{exciton}} \approx \eta \mathrm{N}_a \hbar\omega \quad (4)$$

where is $\hbar\omega$ - photon energy (laser wavelength), N_a - number of absorbed photons (dependent on fluence and exposure time of laser), η<1 - energy-transfer factor (losses to photons, heat, defects). Considering the fact, that decoupling begins when $\eta \mathrm{N}_a \hbar\omega \geq E_{\text{spin-spinsoliton}}$, the threshold laser energy is given by eq. (5):

$$\hbar\omega_{\text{th}} \sim \frac{E_{\text{spin-spinsoliton}}}{\eta \mathrm{N}_a} \quad (5)$$

Decoupling of exciton and spin-spin solitons is expected when the accumulated exciton soliton energy (determined by photon energy, laser fluence, and exposure time), becomes comparable to, or exceeds, the characteristic energy threshold for soliton decoupling. Higher photon energies reduce the required exposure and fluence to reach this regime.

Since laser excitation generates exciton solitons in the α-Fe_2O_3 precursor (which are coupled to spin–spin solitons), it is considered that the energy, transferred through this coupled soliton system, represents additional parameter, by which the absorbed laser energy can influence the magnetic state of the material during formation of the ε-Fe_2O_3 phase. For this reason, given study proposes that the calculated soliton energy and decoupling threshold can be used as indicators of the energy-transfer process, and its potential influence on the magnetic ordering of the resulting ε-Fe_2O_3 phase.

Although an analytical threshold condition for soliton coupling is derived, the effective coupling is dependent on laser photon energy, laser fluence, exposure time (number of pulses), material-dependent coupling strength, and dissipative energy-loss channels. Therefore, the obtained analytical result could be further probed using numerical simulations.

To perform numerical application of the postulated theoretical basis, here is important to notice that numerical computational basis of the model is relied on the probing of the laser excitation threshold, sufficient to induce decoupling of the exciton and spin-spin solitons. Afterwards, two complementary phenomenological descriptions (representing two complementary parameterizations, probing relation between different synthesis parameters and normalized magnetization (M_{norm})) were introduced. First parameterization approach investigate the response of the Mnorm to photon energy, and other controllable parameters. Second parameterization approach (representing PLD cumulative-energy formulation) enable monitoring of the impact of the accumulated laser energy, relative to the effective threshold, on the M_{norm}. Different parameterization formulations relies on various phenomenological expressions for M_{norm}, introduced in order to consider different aspects of M_{norm} behavior.

2.1. Numerical methods for laser-induced decoupling and magnetization estimation

To investigate the effect of laser irradiation on the stability of coupled exciton–spin solitons in α-Fe_2O_3 and the resulting magnetization of the synthesized ε-Fe_2O_3 phase, a numerical solver was developed in MATLAB [16]. This solver is based on the analytical expressions derived for the total soliton energy and the critical energy required for decoupling. It allows systematic variation of the key parameters influencing soliton coupling, including: photon energy of the laser, effective laser exposure time (t_{exp}), or energy-transfer efficiency factor (η), describing the fraction of the relevant absorbed photon energy transferred to the exciton-soliton excitation.

For each set of parameters, the algorithm evaluates whether the absorbed laser energy exceeds the threshold for decoupling the coupled soliton state. If the threshold is surpassed, the program estimates the corresponding reduction in M_{norm}, based on the proportionality between spin–spin soliton energy and the quantized magnetization.

All input parameters are constrained to experimentally relevant ranges, and energy calculations follow the semi-classical quantization framework outlined in Ref. [12] (eq. (9)–(11)).

This approach provides a computationally efficient method to assess the influence of laser parameters on the magnetic properties of ε-Fe_2O_3, without requiring full time-dependent PLD simulations. It also serves as a predictive tool for optimizing synthesis conditions in experiments.
Further, to perform numerical magnetization modeling, in order to quantify the effect of laser-induced soliton decoupling on the magnetization of ε-Fe_2O_3 thin films, a MATLAB-based Magnetization Calculator was developed.

A MATLAB-based solver and Magnetization Calculator were developed to evaluate the analytical threshold relations and the resulting normalized magnetization. The program systematically varies photon energy, exposure time, energy-transfer efficiency, coupling strength, damping, and soliton amplitude, allowing individual and combined parameter effects to be examined without full time-dependent PLD simulations.

3. RESULTS AND DISCUSSION

To visualize the threshold photon energy for decoupling of the exciton-spin-spin soliton system, chosen parameters were systematically varied, such as: spin-spin soliton energy ($E_{\text{spin-spin soliton}}$ [eV]), number of adsorbed photons (N_a), laser fluence [J/cm^2], and exposure time [s]. These variations were modeled, using eq. (6):

$$N_a = \frac{F_p A_{ill}(1-R) A_{absorp} f_{rep} t_{exp}}{E_{ref}} = F_p t_{exp} \alpha_{eff} \qquad (6)$$

where F_p is the laser fluence per pulse, t_{exp} is the total irradiation/deposition duration corresponding to the pulse number, N_{pulses} ($N_{pulses} = f_{rep} t_{exp}$), A_{ill} – illuminated area, R – reflectivity, A_{absorp} – absorptance, f_{rep} – pulse repetition rate, and E_{ref} is reference photon energy. Accordingly, eq. (6) represents a phenomenological estimate of the effective number of absorped photon excitations, during PLD. The effective photon-absorption factor, α_{eff}, incorporates the illuminated area, optical losses, target absorptance, pulse repetition rate, and a reference photon-energy scale; its value is treated as an adjustable model parameter ($\alpha_{eff} = 10^{18} cm^2 J^{-1} s^{-1}$).

The numerical simulations yield the results shown in Fig. 1 and Fig. 2. Since eq. (3) is phenomenological energy-coupling relation, Fig. 1 and Fig. 2 appears as a result of phenomenological threshold/decoupling calculation, providing deeper insight in the conditions, under which laser excitation become sufficient for decoupling. Fig. 1 presents the dependence of the threshold photon energy on laser fluence, for different energy-transfer efficiencies at a fixed exposure time.

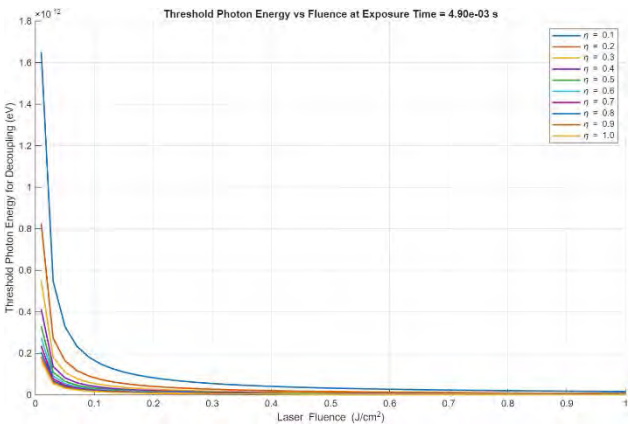


Figure 1. Dependence of threshold photon energy on laser fluence at a fixed exposure time (4.90×10^{-3} s) for various energy-transfer factors (η).

Each curve represents the threshold photon energy required to initiate soliton decoupling as a function of fluence, with different colors, corresponding to different values of η. The plot illustrates impact of the increasing fluence on the lowering of the threshold energy, and highlighting the nonlinear interplay between laser parameters and soliton stability. Fig. 1 shows an inverse dependence of threshold photon energy on fluence: increasing fluence increases the number of absorbed photons and therefore lowers the photon energy required for decoupling. The convergence of the curves at higher fluence indicates diminishing sensitivity to further fluence increases.

Fig. 2 provides a complementary view through a three-dimensional color map, illustrating the variation of the threshold photon energy with both fluence and exposure time, for a fixed energy-transfer factor: η=0.5.

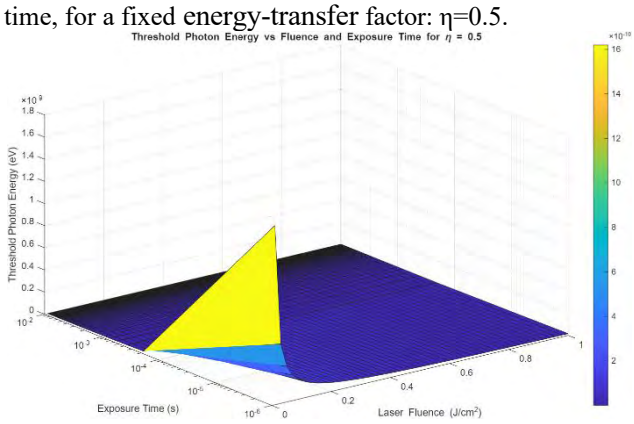


Figure 2. Dependence of threshold photon energy on fluence and exposure time, for a fixed energy-transfer factor: η=0.5.

Fig. 2 shows the corresponding dependence on fluence and exposure time. At low fluence and short exposure, higher photon energies are required for decoupling, whereas increasing either parameter lowers the threshold because of cumulative energy absorption. The surface gradually approaches saturation, indicating diminishing sensitivity at higher fluence and exposure. These trends are consistent with the analytical behavior of the model reported in Ref. [12].

To directly link soliton decoupling to magnetic response, a

Matlab-based Magnetization Calculator was employed. This tool computes the total soliton energy, evaluates the decoupling condition, and outputs the normalized magnetization as a function of laser photon energy and additional synthesis parameters. For this purpose, it is introduced the simplified phenomenological expression (eq. (7)) for evaluation of the M_{norm}, in order to facile interactive visualization (shown in Fig. 3). Thus, eq. (7) represents phenomenological parameter-sensitivity response of M_{norm}, proposed considering first parameterization approach:

$$M_{\text{norm}} = \exp\left(-\gamma \frac{E_{\text{threshold}}}{E_{\text{photon}}} \frac{t_{\text{ref}}}{t_{\text{pulse}}} \frac{T_{\text{ref}}}{T_{\text{target}}} \frac{\rho_{\text{ref}}}{\rho_0}\right) \quad (7)$$

where $E_{threshold}$ denotes the laser energy required to initiate exciton–spin soliton decoupling, as determined from the analytical model. The quantities t_{ref}, T_{ref}, and ρ_{ref} are reference values introduced for normalization, ensuring that the resulting magnetization is expressed as a dimensionless quantity between 0 and 1. The parameter γ represents a dimensionless phenomenological damping parameter ($0 \leq \gamma \leq 1$), with larger values corresponding to stronger damping. The damping coefficient is introduced to account for energy-loss mechanisms present during laser irradiation and soliton propagation, including phonon scattering, thermal dissipation, lattice defects, and non-radiative recombination. These processes reduce the fraction of absorbed laser energy that effectively contributes to soliton coupling and magnetic ordering, making γ an essential parameter for realistic modeling of experimental conditions. Consequently, eq. (7) enables estimation of the effective M_{norm}, incorporating both energy input and dissipative effects.

Higher photon energy, longer pulse duration, lower damping, or higher soliton amplitude, result in increased M_{norm}, reflecting more efficient energy transfer and stronger magnetic ordering in the synthesized thin films.

The resulting dependence of M_{norm} on photon energy is shown in Fig. 3. Accordingly, Fig. 3 shows response of the M_{norm} on the photon energy, and other controllable parameters.

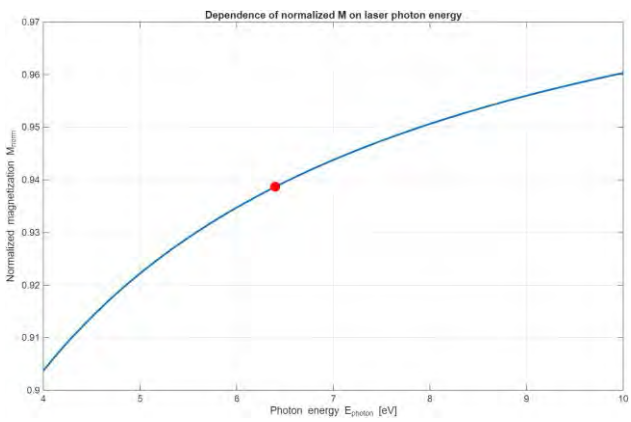


Figure 3. Dependence of normalized magnetization on laser photon energy.

Fig. 3 demonstrates a pronounced threshold-like behavior: at low photon energies, the normalized magnetization remains low, indicating insufficient energy to sustain exciton–spin–spin soliton coupling. As photon energy increases, the magnetization rises steeply, approaching unity when the photon energy reaches or exceeds the coupling threshold. This confirms that higher-energy photons more efficiently preserve magnetic ordering during the phase transformation from α-Fe_2O_3 to ε-Fe_2O_3. The red marker in Fig. 3 corresponds to a specific parameter set selected interactively via Matlab Live Script sliders. Its vertical position reflects the influence of additional parameters—such as pulse duration, soliton amplitude, temperature, and damping—demonstrating that even at high photon energies, non-optimal values of these parameters can significantly reduce the resulting magnetization. This emphasizes the necessity of multi-parameter optimization, rather than reliance on photon energy alone.

Noteworthy, the PLD Magnetization Calculator enables calculation of M_{norm}, and a simultaneous assessment of multiple parameters affecting the normalized magnetization, with results summarized in tabular form (Table 1). Since Table 1 results are computed, relying on the second parameterization approach, for numerical computations are applied next eq. (8-10):

$$E_{laser} = E_{photon}\eta ln(1 + N_{pulses}) \quad (8)$$

$$E_{total} = E_{soliton} + E_{laser} \quad (9)$$

$$E_{threshold} = E_{bond} + E_{thermal} \quad (10)$$

which is resulting in the next phenomenological expression, describing M_{norm} response: $M_{\text{norm}} = \frac{1}{1+(E_{total}/E_{threshold}-1)^2}$ (11). The critical photon energy is obtained, considering condition that: $E_{total} = E_{threshold}$, resulting in the eq. (12):

$$E_{critical} = \frac{E_{threshold} - E_{soliton}}{\eta ln(1+N_{pulses})} \quad (12).$$

The calculations are performed, considering values, used as computational cases spanning a realistic deposition-time scale. In PLD synthesis, the total irradiation time is related to the number of applied laser pulses through: $N_{pulses} = f_{rep}t_{exp}$, while the pulse duration describes the width of each individual laser pulse. To reflect this distinction, the present calculations use a repetition rate of 5 Hz and consider representative pulse numbers of 5,000–20,000, corresponding to total irradiation times of 1,000–4,000 s (approximately 17–67 min), while a pulse duration of 20 ns is used as a representative PLD pulse width. The selected pulse numbers are representative computational cases, rather than experimentally reported pulse counts for a specific ε-Fe_2O_3 film. Numerical results are represented in Table 1.

Table 1. Calculated M_{norm} for representative PLD conditions, using KrF, ArF, and F_2 laser excitation.

Laser	Photon Energy [eV]	Num Pulses	Exposure [s]	Total energy	Threshold energy	Critical energy	M_{norm}
KrF	4.9	5000	1000	34.888	1.5	0	0.0020143
KrF	4.9	10000	2000	37.605	1.5	0	0.0017231
KrF	4.9	20000	4000	40.322	1.5	0	0.0014907
ArF	6.4	5000	1000	45.109	1.5	0	0.0011817
ArF	6.4	10000	2000	48.657	1.5	0	0.0010107
ArF	6.4	20000	4000	52.206	1.5	0	0.00087434

F_2	8.1	5000	1000	56.693	1.5	0	0.00073807
F_2	8.1	10000	2000	61.184	1.5	0	0.00063124
F_2	8.1	20000	4000	65.675	1.5	0	0.00054603

This tabular output facilitates rapid comparison between different laser sources and exposure conditions, offering a practical and predictive framework for optimizing PLD synthesis strategies. Table 1 proposes impact of the accumulated laser energy, relative to the effective threshold, on the M_{norm}.

The motivation for this calculator is based on the fact that published PLD studies roughly establish that laser and deposition conditions influence ε-Fe_2O_3 phase formation and magnetic properties, while the presented model explores the sensitivity of M_{norm} to photon energy.

Literature search showed that temperature and structural/growth parameters demonstrably affect the measured magnetization [17]. Also, Corbellini et al. discussed the relevance of photon energy, pulse duration, repetition rate, fluence and exposure/pulse number, as PLD parameters [1]. They listed the experimental PLD parameters explicitly: KrF, 248 nm, ~22 ns pulse duration, 5 Hz repetition rate, and ~1.8 J/cm² fluence, for the films, characterized magnetically by VSM, what demonstrated that PLD processing parameters and the resulting magnetic properties are experimentally connected, which supports the general motivation for our parameterized calculator. In 2022 is published study, where authors reported that oxygen pressure, deposition/annealing temperature, and laser beam energy significantly affect epitaxial growth of ε-Fe_2O_3; their ~20 nm films show strong magnetic anisotropy and a coercivity of about 600 Oe [18]. This is very close to what our calculator is trying to represent. Authors explicitly claimed that laser beam energy affects the epitaxial growth of ε-Fe_2O_3. That is very supporting for Fig. 3, represented in this study, because our exponential model specifically explores photon energy. We therefore have literature support for investigating laser-energy dependence, even though the literature does not validate our particular exponential formula.

Albeit, the calculator is not quantitatively calibrated against experimental values. Published PLD studies demonstrate that the magnetic properties of ε-Fe_2O_3 thin films depend on synthesis conditions, including temperature, film thickness, and laser/deposition parameters. The present model therefore provides a phenomenological parameter-sensitivity analysis, rather than an absolute prediction of magnetization.

4. CONCLUSION

In this work, a combined analytical and numerical framework was developed to investigate laser-induced exciton–spin–spin soliton interactions during PLD synthesis of ε-Fe_2O_3 thin films and their impact on magnetic properties. Building upon a previously proposed soliton-based model, analytical expressions were derived to describe energy transfer, soliton coupling, and decoupling thresholds as functions of laser photon energy, fluence, exposure time, and coupling strength.
Numerical basis of the model is given by eq. (6), which enables deeper insight in the amount of photon energy, required for decoupling, and illustrates sensitivity of the number of absorbed photons variations, on the parameters, such as fluence, exposure and coupling efficiency. Expressions for M_{norm} (eq. (7) and (11)) are obtained by two complementary phenomenological descriptions: first one provided a normalized parametric response for interactive exploration of photon energy, pulse duration, temperature, soliton amplitude and damping, and second approach is based on cumulative laser-induced energy relative to a decoupling threshold.

Numerical simulations revealed a pronounced nonlinear dependence of the threshold photon energy on laser fluence and exposure time. The resulting two- and three-dimensional representations demonstrated characteristic saturation behavior, showing that cumulative energy absorption can significantly reduce the photon energy required for soliton decoupling. These trends are fully consistent with the analytical model, validating its ability to capture the essential physics of laser–matter interactions governing phase transformation and magnetic ordering.

To quantify the magnetic consequences of soliton decoupling, a MATLAB-based Magnetization Calculator was developed. By incorporating additional physically relevant parameters—such as soliton amplitude, coupling strength, and damping effects—the tool enables estimation of normalized magnetization under realistic experimental conditions. The simulations illustrate that the dependence of normalized magnetization on photon energy depends on the adopted phenomenological parameterization, while non-optimal combinations of exposure time, damping, or coupling strength can substantially reduce the magnetic response, underscoring the importance of multi-parameter optimization.

Overall, the presented approach provides a computationally efficient methodology for exploring PLD parameter space, without the need for full time-dependent PLD simulations. The combined framework provides a computationally efficient approach for exploring the sensitivity of soliton coupling and normalized magnetization to PLD parameters. The methodology can be extended to related oxide systems in which laser-induced energy transfer influences phase stability and functional properties.

Acknowledgments

The research was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (451-03-66/2024-03/200017).

References

[1] CORBELLINI, L., LACROIX, C., HARNAGEA, C., KORINEK, A., BOTTON, G. A., MÉNARD, D., PIGNOLET, A.: Epitaxially stabilized thin films of ε-Fe_2O_3 (001) grown on YSZ (100), *Scientific Reports*, **7** (2017) 3712.

[2] JEONG, J. H., KIM, H. H., LEE, J.-G., BAEK, Y.: Epsilon iron oxide (ε-Fe_2O_3) as an electromagnetic functional material: Properties, synthesis, and applications, *Journal of Powder Materials*, **31**(6) (2024) 465–479.

[3] GORBACHEV, E., SOSHNIKOV, M., WU, M., ALYABYEVA, L., MYAKISHEV, D., KOZLYAKOVA, E., LEBEDEV, V., ANOKHIN, E., GORSHUNOV, B., BRYLEV, O., KAZIN, P., TRUSOV, L.: Tuning the particle size, natural ferromagnetic resonance frequency and magnetic properties of ε-Fe_2O_3 nanoparticles prepared by a rapid sol–gel method, Journal of Materials Chemistry C, 9 (2021) 6173–6181.

[4] TOKORO, H., NAMAI, A., OHKOSHI, S.: Advances in magnetic films of epsilon-iron oxide toward next-generation high-density recording media, Dalton Transactions, 50(2) (2021) 452–459.

[5] SHAO, Z., REN, S.: Rare-earth-free magnetically hard ferrous materials, Nanoscale Advances, 2 (2020) 4341–4349.

[6] VU QUOC VIET, SALAWU, Y. A., SON, W. H., RHYEE, J.-S., LEE, N.-S., KIM, H.-J.: Bi-stability and orientation change of a thin α-Fe_2O_3 layer on an ε-Fe_2O_3 (004) surface, ACS Omega, 4 (2019) 10489–10497.

[7] HUFF, M.: Residual stresses in deposited thin-film material layers for micro- and nano-systems manufacturing, Micromachines, 13(12) (2022) 2084.

[8] OHKOSHI, S., TOKORO, H.: Hard magnetic ferrite: ε-Fe_2O_3 free, Bulletin of the Chemical Society of Japan, 86(8) (2013) 897–907.

[9] DVORTSOVA, E., SUTURIN, S.: Technological peculiarities of epsilon ferrite epitaxial stabilization by pulsed laser deposition, Molecules, 5(4) (2022) 32.

[10] CHEN, L., WANG, X., ZHANG, Y., LI, H., XU, Q.: Microstructure and physical properties of ε-Fe_2O_3 thin films fabricated by pulsed laser deposition, Journal of Applied Physics, 131(4) (2022) 045301.

[11] YU, S., ZHANG, P., LI, J.: Stoichiometry, phase, and texture evolution in PLD-grown hexagonal barium ferrite films, arXiv preprint arXiv:1904.03157 (2019).

[12] NIKOLIĆ, V., MARIANO, J. F. M. L.: Optimisation of PLD synthesis of ε-Fe_2O_3 thin films, in: International Scientific Conference on Military Sciences – VojNA 2025, Belgrade, Serbia, 11–12 September 2025, pp. 1–6.

[13] GU, C. H., HU, H. S., ZHOU, Z. X.: Soliton Theory and Its Applications, Springer, Berlin, 1995.

[14] ABORINGONG, E. N. N., DIKANDÉ, A. M.: Soliton lattices originating from excitons interacting with high-intensity fields in finite molecular crystals, European Physical Journal Plus, 134 (2019) 1–11.

[15] STOYCHEV, K. T., PRIMATAROWA, M. T., MARINOV, K.: Exciton-polariton solitary waves, European Physical Journal B, 29 (2002) 301–304.

[16] MATHWORKS, INC.: MATLAB, version R2024a, The MathWorks Inc., Natick, Massachusetts, USA, 2024.

[17] Suturin SM, Korovin AM, Sitnikova AA, Kirilenko DA, Volkov MP, Dvortsova PA, Ukleev VA, Tabuchi M, Sokolov NS. Correlation between crystal structure and magnetism in PLD grown epitaxial films of ε-Fe2O3 on GaN. Sci Technol Adv Mater. 2021 Feb 4;22(1):85-99.

[18] Chen, Shanshan, Yixiao Jiang, Tingting Yao, Ang Tao, Xuexi Yan, Fang Liu, Chunlin Chen, Xiuliang Ma, and Hengqiang Ye. "Microstructure and physical properties of ε-Fe2O3 thin films fabricated by pulsed laser deposition." Micron 163 (2022): 103359

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

FLOQUET-ENHANCED ELECTRON TRANSPORT IN 1D Fe–X MAGNETIC MOLECULAR SPINTRONIC CHAINS

JOSE F. M. L. MARIANO
Universidade do Algarve, FCT, Campus de Gambelas, 8005-139 Faro, Portugal
Center of Physics and Engineering of Advanced Materials (CeFEMA), IST, Universidade de Lisboa, Av. Rovisco Pais, 1096-001 Lisboa, Portugal, jmariano@ualgt.pt

VIOLETA N. NIKOLIĆ
Vinča Institute of Nuclear Sciences „Vinča", University of Belgrade, Mike Petrovića Alasa 12-14, P.O. Box 522, 11001 Belgrade, Serbia, violeta@vin.bg.ac.rs

Abstract: *In this study is theoretically investigated light-driven electron transport in one-dimensional Fe–X molecular chains (X = Fe, Ni, Co) by combining Floquet band analysis with open-system transport modeling—the first framework to directly connect quasi-energy spectra under periodic driving with dissipative electron transport. Floquet calculations reveal that symmetry, orbital hybridization, and exchange splitting govern photon-assisted band mixing, while transport simulations quantify spin-resolved current enhancement. Results showed that Fe–Co motifs maximize photon-assisted conduction due to strong hybridization and high mid-band photon weight, Fe–Ni exhibits moderate enhancement, and Fe–Fe shows the least. This unified approach provides predictive guidelines for designing molecular spintronic devices optimized for light-controlled, high-efficiency electron transport.*

Keywords: *Molecular spintronics, Floquet bands, Photon-assisted transport, 1D magnetic Fe–X chains*

1. INTRODUCTION

Nowadays, molecular spintronics has attracted significant interest from the scientific community due to its potential to enable nanoscale devices with enhanced performance and novel functionalities [1]. This interdisciplinary field lies at the intersection of organic spintronics, molecular magnetism, and molecular quantum technologies. Molecular spintronic devices exploit the electron spin for data processing and storage, with applications ranging from ultrafast and energy-efficient computing to high-sensitivity sensing, which could be used in variety of potential military applications [2-5].

In this study, we combine Floquet band structure analysis (Model I) with open-system transport modeling (Model II) to investigate 1D Fe–X molecular chains, where X represents magnetic atoms (Fe, Ni, Co). The two postulated models are complementary: Model I provides the periodic-system Floquet electronic structure, while Model II uses the corresponding Fe–X motif, as an effective molecular junction to investigate photon-assisted transport. As a result, it was possible to quantify the influence of orbital hybridization and periodic driving on electron transport via photon-assisted resonances. Based on our simulations, we obtained deeper insight in the efficient electron conduction, and highlight the design principles for photon-assisted transport, which could facile optimization of the light-driven transport in molecular spintronic devices.

2. THEORETICAL MODEL AND NUMERICAL METHODS

Two complementary theoretical models are employed and combined, with the aim to investigate light-driven electron transport in a 1D Fe–X molecular system. Model I describes a periodically driven tight-binding Fe–X chain (where X denotes Fe, Ni, and Co), and it is used to analyze Floquet band formation and hybridization effects.
Model I, denoting to the periodically driven tight-binding Fe–X chain, is relied on the static tight-binding Hamiltonian, constructed by considering a 1D periodic chain, composed of alternating Fe and X sites, within a single unit cell.

It is important to note that we begin calculations from a tight-binding Hamiltonian formulated in a localized orbital basis, where on-site energies describe Fe and X atomic orbitals and hopping terms, represent nearest-neighbor coupling and hybridization. This Hamiltonian initially written in real space. For the periodic one-dimensional Fe–X chain considered in the band-structure analysis, translational symmetry allows the application of Bloch's theorem. By performing a Fourier transform to momentum space, the real-space tight-binding Hamiltonian is recast into a k-dependent Bloch Hamiltonian H(k). The Bloch Hamiltonian is therefore not a different physical model, but rather the momentum-space representation of the same tight-binding Hamiltonian under periodic boundary conditions. This representation enables calculation of the band structure and, subsequently, its Floquet extension under time-periodic driving.

Corresponding author: V. Nikolić;
E-mail: violeta@vin.bg.ac.rs

DOI: 10.5937/OTEH26104M

Consequently, discussed Model I provided band-structure level understanding of orbital hybridization, Floquet sideband formation, resonance conditions. To perform full Matlab numerical simulating, Model I was considered in combination with Model II. In order to discuss electron transport under applied bias, an open quantum system, representing an effective Fe–X dimer, coupled to electronic reservoirs, was considered. For the quantitative transport analysis, the effective Fe–X dimer is coupled to left and right electronic reservoirs within a Floquet nonequilibrium Green's-function (Floquet-NEGF) formulation. This allows the bias-dependent driven and static currents to be evaluated consistently, with the same Fe–X parameter sets used in Model I.

Numerical procedure, performed in order to apply Model I on a considered system, was based on the construction of $H_0(k)$, for each value of k. Afterwards, it was builded truncated Floquet matrix, which diagonalization was performed numerically, and quasi-energies and orbital weights were extracted, for each vlaue of k. All calculations were performed, using standard matrix diagonalization routines in MATLAB (R2024a) [8].

Below is given theoretical postulation of the model, relying on two complementary modeling approaches. In order to construct appropriate static tight-binding Hamiltonians, few model premises were introduced. In the study is considered 1D periodic chain, consisted of motif, containing two atoms per unit cell (except Fe-Fe symmetric case): Fe atom, and X-atom (where X denotes to Ni, and Co atom). Model is based on nearest-neighbor hopping, and inter-site hybridization, as well as spin splitting is also considered, since discussed atoms are magnetic.

Considered Bloch Hamiltonians, postulated for appropriate 1D Fe-X molecular chains, are shown below. In the case of Fe-Fe chain, Bloch Hamiltonians is simplified, since identical atoms are considered, eq. (1):

$$H_0^{FeFe}(k) = \begin{pmatrix} \varepsilon_{Fe} & te^{-ika} \\ te^{-ika} & \varepsilon_{Fe} \end{pmatrix} \quad (1)$$

where ε_{Fe} denotes to on-site energy, t is the nearest-neighbor hopping amplitude, and a is the lattice constant. Introduction of magnetism in the model is mathematically achieved by introducing exchange splitting: $\varepsilon_{Fe,\uparrow} = \varepsilon_{Fe} + \frac{\Delta_{ex}}{2}, \varepsilon_{Fe,\downarrow} = \varepsilon_{Fe} - \frac{\Delta_{ex}}{2}$, while for the heteronuclear chains, the same construction is applied separately to Fe and X: ↑ ↓

$$\varepsilon_{i,\uparrow/\downarrow} = \varepsilon_i \pm \frac{\Delta_i}{2}; i = Fe, X \quad (2)$$

where is Δex – term of magnetic exchange splitting. Here should be mentioned that, in the case of Fe-Fe unit cell, it was considered large value of Δex, while it is intermediate for the case of Fe-Co dimer, and finally, the smallest is for Fe-Ni dimer. Performed mathematical step provided Hamiltonian, which is block-diagonal in spin.

Accordingly, the Hamiltonian doubles in spin space (4x4 per k). Because the two sites are parametrized identically in the Fe–Fe case, the static Hamiltonian has the corresponding site-exchange symmetry and does not contain onsite-energy asymmetry.

Bloch Hamiltonian of 1D Fe-Ni molecular chain is given by eq. (3):

$$H_0^{FeNi}(k) = \begin{pmatrix} \varepsilon_{Fe} & V_{FeNi} + te^{-ika} \\ V_{FeNi} + te^{-ika} & \varepsilon \end{pmatrix} \quad (3)$$

where ε_{Fe} and ε_{Ni} represent H on-site 0 (k) = energies, t is the hoppling amplitude, V represents the Fe-Ni hybridization strength, a is the lattice constant, k is the crystal momentum. ε_{Ni} is characterized by slightly lower value compared to ε_{Fe}, since Ni is characterized by higher number of electrons in d-band, although both atoms belongs to 3d metals, which implies moderate hybridization.

Finally, Bloch Hamiltonian of 1D Fe-Co molecular chain, eq. (4):

$$H_0^{FeCo}(k) = \begin{pmatrix} \varepsilon_{Fe} & V_{FeCo} + te^{-ika} \\ V_{FeCo} + te^{-ika} & \varepsilon_{Co} \end{pmatrix} \quad (4)$$

where ε_{Fe} and ε_{Co} represent H on-site 0 (k) = energies, t is the hoppling amplitude, V represents the Fe-Co hybridization strength, a is the lattice constant, k is the crystal momentum. Since both atoms, Fe and Co, are characterized by very similar in d-band character, it is considered smaller onsite asymmetry than Fe-Ni, which consequently could be characterized by strong exchange splitting.

In order to perform numerical methods, below are given Tables, giving set of parameters for characteristic 1D Fe-X motifs. The parameter sets were selected as representative values consistent with the distinct electronic coupling, level alignment, and magnetic characteristic of the Fe-Fe, Fe-Ni, and Fe-Co configurations, reported in the literature [6, 7]. They are therefore used as model parameterizations, rather than material-specific fitted parameters. Parameter values used in the simulations are showed in Table 1-3.

Table 1. Parameters used in all numerical simulations of the 1D Fe–Fe molecular chain.

Parameter	Symbol	Typical value
Hopping amplitude	t	1.2 eV
On-site energy	ε_{Fe}	0.0 eV
Magnetic splitting exchange constant	Δ_{ex}	1.5 eV
Driving amplitude	A	0 -3 eV
Driving frequency	ω	1 – 3 eV

Given parameters are chosen, considering the fact that discussed band is metallic, and characterized by the absence of intrinsic gap, while Floquet gap opens only via driving.

Table 2. Parameters used in all numerical simulations of the 1D Fe–Ni molecular chain.

Parameter	Symbol	Typical value
Hopping amplitude	t	1.0 eV
Fe-Ni orbital hybridization	V	1.5 eV
On-site energy	ε_{Fe}	0.2 eV
On-site energy	ε_{Ni}	- 0.3 eV
Driving amplitude	A	0 -1 eV
Driving frequency	ω	1 – 3 eV

Based on chosen parameters in Table 2, it is expected appearance of moderate bonding-antibonding splitting, and asymmetric orbital weights.

Table 3. Parameters used in all numerical simulations of the 1D Fe–Co molecular chain.

Parameter	Symbol	Typical value
Hopping amplitude	t	1.1 eV
Fe-Co orbital hybridization	V	1.8 eV
On-site energy	ε_{Fe}	0.1 eV
On-site energy	ε_{Co}	- 0.1 eV
Magnetic splitting exchange constant	Δ_{ex}	1.2 eV
Driving amplitude	A	0 -1 eV
Driving frequency	ω	1 – 3 eV

Based on chosen parameters in Table 3, it is expected appearance of strong hybridization, and large exchange splitting.

Further, under periodic driving with field amplitude A and frequency ω, it is considered the time-dependent Hamiltonian. In order to incorporate time-dependent driving in a postulated theoretical model, it was used modulation given in eq. (5):

$$H(k,t) = H_0(k) + A\cos(\omega t)\hat{O} \quad (5)$$

where $\hat{O}$ is the operator coupled to the external field. Consequently, driving modulated inter-site coupling by expression: V → V + Acos(ωt).

In order to collect spin-resolved spectra by numerical simulations, it was performed Floquet extension in harmonic space: m = -N_F, …, + N_F, where N_F = 1 depicts to the 3 harmonics and 4 spin-orbital states. A first-order Floquet truncation, m = 1, 0, +1, is adopted to retain the leading photon-assisted processes, while keeping the model minimal and transparent. Higher-order multiphoton processes are not considered in the present study. So, total Floquet dimension was 4 x 3 = 12. Accordingly, Floquet matrix is constructed by using next parameters: m = -1, 0, +1, and N_F = 1, as well as 12 x 12 matrix size (for including of the spin impact). The Floquet matrix element (in Block Hamiltonian form) is given by eq. (6):

$$(H_F)_{mn} = H_{m-n}(k) + n\omega\delta_{mn} \quad (6)$$

where is $H_{\pm 1} = \frac{A}{2}\hat{O}$.

Consequently, the resulting matrix is 12 x 12, and it is diagonalized numerically, in order to obtain quasi-energies ε_α (k), spin-resolved orbital weights, and photon weights.
To combine Model I with Model II, the effective Fe–X dimer is coupled to left and right noninteracting electronic reservoirs. The transport calculation is performed within a Floquet nonequilibrium Green's-function (Floquet-NEGF) formulation. The lead Hamiltonian is given by eq. (8):

$H_{leads} = \sum \varepsilon_{k,\alpha}\, c_{k\alpha}{}^{\dagger} c_{k\alpha}$ (8), where α=L, R denotes the left or right reservoir, and k labels the lead states. The coupling to the molecular dimer is described through the corresponding lead self-energies. In the present calculations, symmetric wide-band lead coupling was assumed, with: $\Gamma_L = \Gamma_R$ = 0.20 eV.

The bias-dependent current is obtained from the Floquet transmission function by summing the contributions from the retained Floquet sectors: m = -1, 0, +1 . Both the static (A=0) and periodically driven (A = 0.6 eV) currents are calculated using the same molecular parameters and lead couplings. This provides the quantitative transport comparison presented below.

3. RESULTS AND DISCUSSION

The combined application of Model I (Floquet band engineering) and Model II (nonequilibrium open-system transport) enables a systematic comparison of light-driven transport properties in three magnetic 1D motifs: Fe–Fe, Fe–Ni, and Fe–Co. This complementary approach allows comparison of Floquet band-structure features with the bias-dependent transport response.

Results of numerical simulation are represented in Fig. 1 – 3, for the case of 1D Fe-Fe molecular chain.

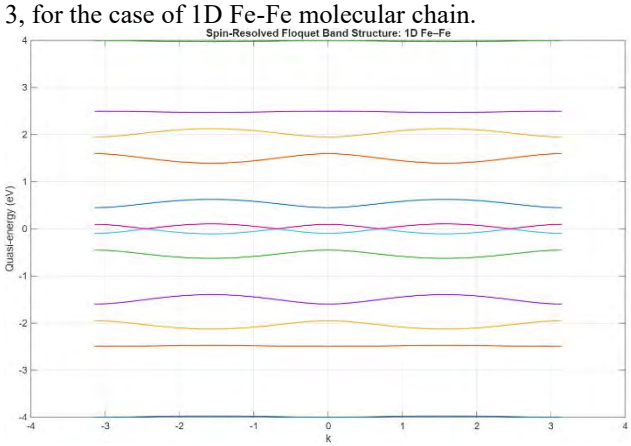


Figure 1. Spin-resolved Floquet quasi-energy band structure of a periodically driven 1D Fe–Fe molecular chain under driving (A = 0.6, ω = 2 eV).

Fig. 1 represents Floquet quasi-energy band structure of the periodically driven 1D Fe–Fe chain, obtained from the k-dependent Bloch Hamiltonian under harmonic driving (the colors of the curves are used for visual clarity to distinguish different bands and do not correspond to a particular contribution). The evolution of averaged (quasi) bands, in a system with a 1D molecular chain, consisting of two sublattices of the same atom, Fe, as a function of the crystal momentum k, in the presence of a time-periodic drive, provided visualization of band hybridization, occurred due to interaction with a photon, in the vicinity of resonance. Since Fe and Fe are characterized by d-band structures but the same electronegativities, it is also characterized by the absence of onsite asymmetry. Symmetric avoided crossings, indicating photon-induced hybridization. Recall that in a band structure, an avoided crossing occurs when two energy bands approach each other as a function of some parameter (like crystal momentum k) but do not cross because the two states are coupled (hybridized). Consequently, spin channels are balanced due to identical onsite energies.

Orbital and photon composition for selected Floquet band for 1D Fe-Fe molecular chain, is given in Fig. 2.

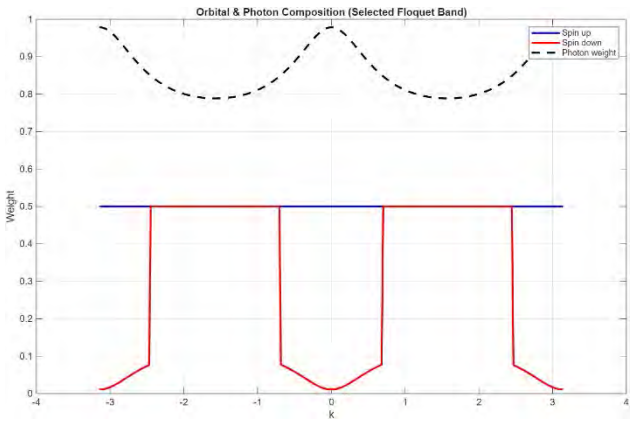


Figure 2. Harmonic composition of a selected Floquet band in Fe–Fe. Blue and red curves represent spin-up and spin-down orbital weights, respectively; black dashed curve shows the total photon weight (sum over non-zero Floquet harmonics).

Fig. 2 represents a visualization of the orbital composition (probability weights of Fe orbitals, differed in spin orientation) of a chosen Floquet band, dependent on the crystal momentum, k. It represents the Floquet band composition across the Brillouin zone, showing the spin-resolved orbital contributions and their variation with crystal momentum. Fig. 2 depict moderate, symmetric photon weight, reflecting uniform Floquet mixing.
In Fig. 3 are represented Floquet quasi-energy spectrum at k=0, for Fe-Fe molecular chain.

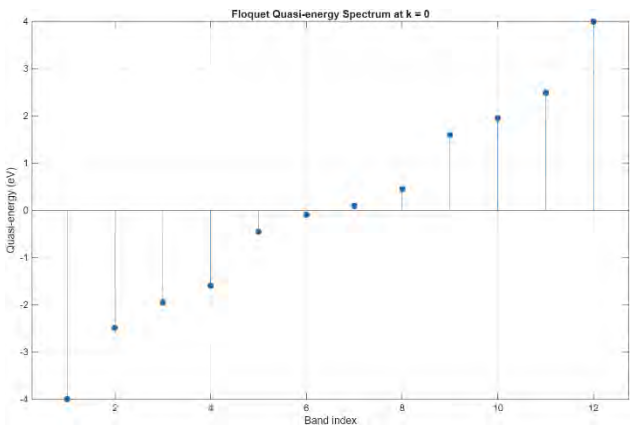


Figure 3. Quasi-energy spectrum at k = 0 for Fe–Fe, showing spin-resolved Floquet levels.

Based on the given dependence of quasi energy on band index, it could be boservable that level splitting indicates photon-mediated coupling. In other words, symmetry reflects identical atoms in the unit cell.

Considering postulated theoretical framework for numerical simulations, Fig. 4 – 6 describes behavior of 1D Fe-Ni molecular chain, modelled within the proposed theoretical frame.

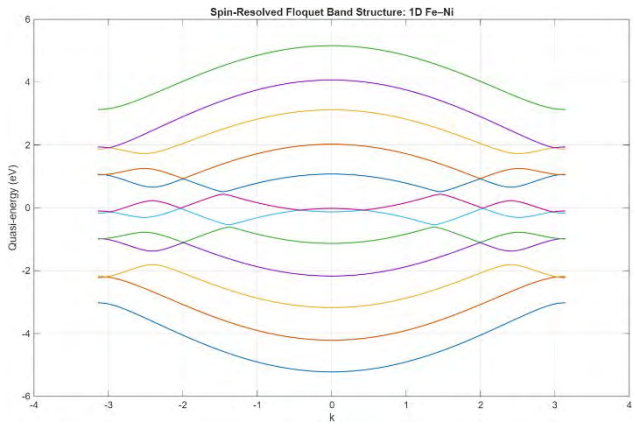


Figure 4. Spin-resolved Floquet quasi-energy band structure of 1D Fe–Ni chain under periodic driving.

Representation of spin-resolved band structure, given in Fig. 4, revealed higher degree of hybridization. Also, observed asymmetry is presented (differed from Fe-Fe chain case), due to intrinsic onsite energy difference, with momentum-dependent avoided crossings reflecting selective photon-assisted hybridization.

In Fig. 5, harmonic composition (Floquet mode weights) of the chosen quasi-energy band, illustrating the mixing of photon sectors induced by periodic driving, is represented (blue curve corresponded to spin up orbital weight, red curve to spin down orbital weight, and black dashed curve to the photon component.

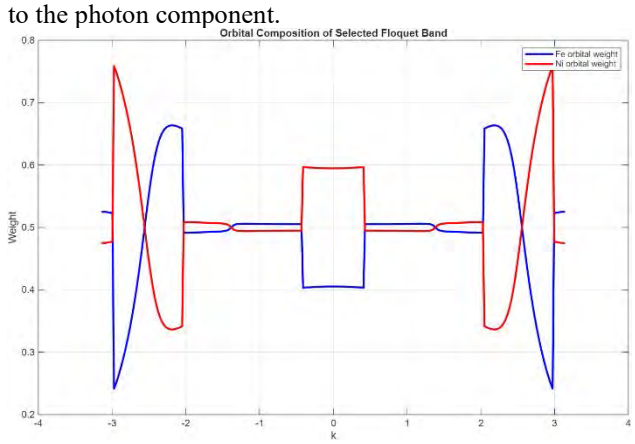


Figure 5. Harmonic composition of a selected Floquet band in Fe–Ni.

Fig. 5 showed that spin-dependent orbital weights (blue/red) highlight asymmetric Floquet mixing and inter-orbital photon coupling.

Further, Fig. 6 described Fe-Ni quasi-energy spectrum at k = 0.

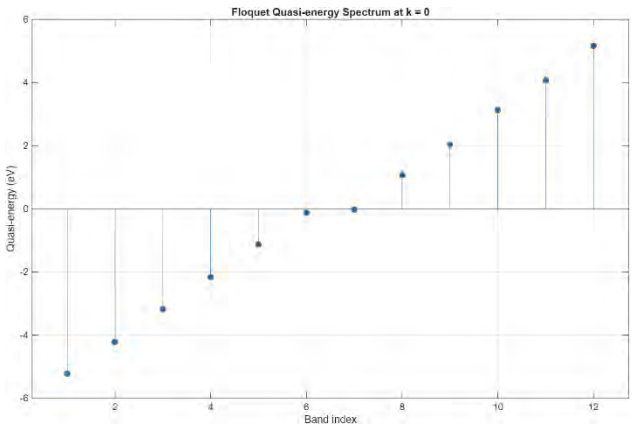


Figure 6. Quasi-energy spectrum at k = 0 for Fe–Ni.

Fig. 6 showed that it is observable uneven level spacing, reflecting intrinsic onsite asymmetry and selective photon-assisted coupling.

Finally, the results of the appropriate simulations, performed for 1D Fe-Co molecular chain, are given in Fig. 7 – 9.

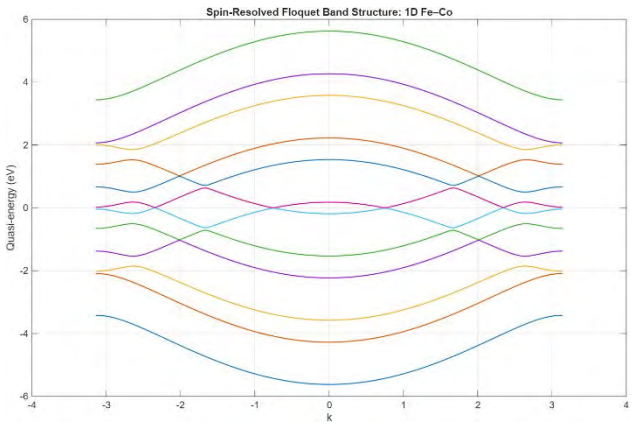


Figure 7. Spin-resolved Floquet band structure of 1D Fe–Co chain under periodic driving.

Differed from Fig. 1 and 4, in Fig. 7 could be observable pronounced avoided crossings and spin-polarized bands indicate strong hybridization and maximal photon-assisted band mixing.

Fig. 8 revealed harmonic composition of a selected Floquet band in Fe–Co. Blue/red curves depict spin-resolved orbital weights.

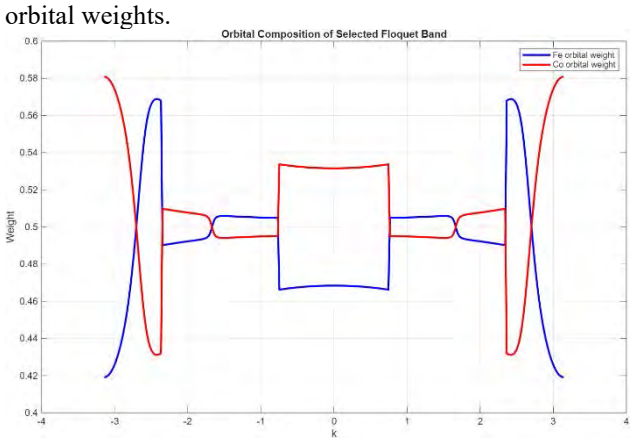


Figure 8. Harmonic composition of a selected Floquet band in Fe–Co. Blue/red curves depict spin-resolved orbital weights.

Finally, in Fig. 9 is represented chosen Fe-Co quasi-energy spectrum at k=0.

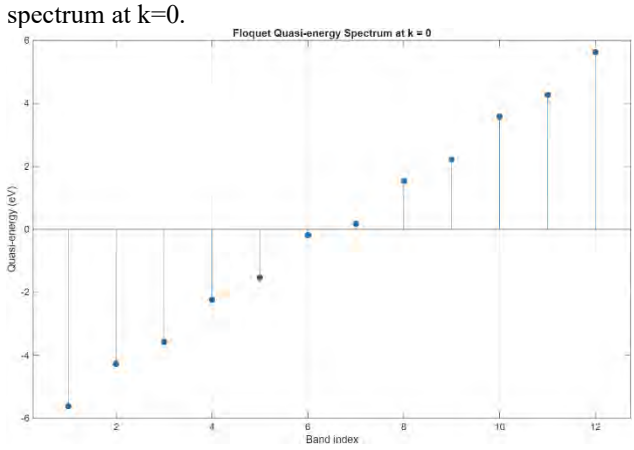


Figure 9. Quasi-energy spectrum at k = 0 for Fe–Co chain.

Fig. 9 showed largest photon-mediated level splitting among the motifs, which is consistent with stronger light-induced inter-orbital coupling within the present model.

Based on performed simulations, it was possible to derive certain conclusions regarding Floquet band structure, and impact of symmetry on hybridization.

The periodically driven Fe–Fe chain exhibits symmetric quasi-energy bands due to identical onsite energies and strong exchange splitting. In the absence of intrinsic onsite asymmetry, Floquet gaps emerge exclusively from photon-induced hybridization at resonance conditions. The resulting avoided crossings are symmetric in momentum space and spin-resolved but balanced. Consequently, light-induced band renormalization is moderate and uniformly distributed across the Brillouin zone.

In contrast, the Fe–Ni chain introduces intrinsic asymmetry ($\varepsilon_{Fe} \neq \varepsilon_{Ni}$), which leads to bonding–antibonding splitting already in the static case. Under periodic driving, this asymmetry produces momentum-dependent Floquet replicas with uneven orbital contributions. Avoided crossings become k-selective and less symmetric, indicating that photon absorption selectively enhances inter-orbital transitions rather than purely intra-sublattice processes.

Within the chosen parameterization, the case of Fe–Co chain exhibits the most pronounced avoided crossings of the orbitals. Due to strong Fe–Co orbital overlap and significant exchange splitting, the Floquet spectrum exhibits multiple pronounced avoided crossings and strong spin-dependent band renormalization. The quasi-energy dispersion reveals enhanced photon-induced coupling particularly in mid-spectrum bands, suggesting optimal resonance matching between driving frequency and intrinsic band separation.

Further, harmonic (photon) composition and Floquet mixing also could be commented. Analysis of harmonic weights reveals qualitative differences in light–matter interaction strength among the three motifs. In Fe–Fe, photon weight remains moderate and nearly symmetric between spin channels, consistent with uniform Floquet mixing. In Fe–Ni, photon weight becomes momentum-dependent and correlates with orbital asymmetry, reflecting selective hybridization between inequivalent

sites. In Fe–Co, photon weight is maximal in central quasi-energy bands and strongly spin-dependent, indicating enhanced light-induced inter-orbital coupling.

The increased photon weight in Fe–Co indicates stronger Floquet mixing associated with the chosen parameterization. The transport calculations further show that this enhanced Floquet mixing can contribute to photon-assisted transport, but does not constitute a universal predictor of current enhancement, which also depends on bias and the open-junction transport conditions.

To probe visual representation of the photon-assisted transport diagnostic for Fe-Fe, Fe-Ni, and Fe-Co pairs, numerical calculations are performed, relying on parameter set values, given in the frame of this model, Fig. 10.

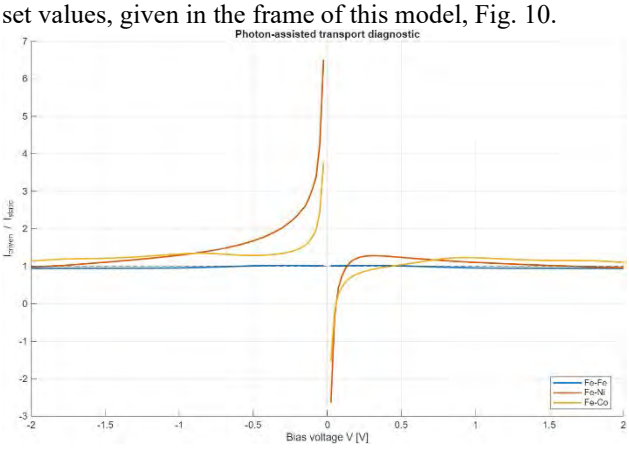


Figure 10. Photon-assisted transport diagnostic for Fe-Fe, Fe-Ni, and Fe-Co.

Fig. 10 is result of visualization of numerical calculations, representing driven/static current ratios, given in Table 4. The full transport calculations were performed over the bias range: $-2 \leq V \leq +2V$. For the statistical comparison in Table 4, the positive-bias interval: 0.1 < V < 2V, was considered, which avoids duplicating the corresponding negative-bias branch and provides a single consistent bias interval for comparing the three systems.

Table 4. Minimum, maximum, and mean driven/static current ratios, over the calculated positive-bias interval: 0.1 < V < 2 V.

Minimum ratio (Fe-Fe)	0.943988
Maximum ratio (Fe-Fe)	1.022696
Mean ratio (Fe-Fe)	0.970489
Minimum ratio (Fe-Ni)	0.952467
Maximum ratio (Fe-Ni)	1.282178
Mean ratio (Fe-Ni)	1.101671
Minimum ratio (Fe-Co)	0.587130
Maximum ratio (Fe-Co)	1.231292
Mean ratio (Fe-Co)	1.107227

Table 4 shows that the photon-assisted response is bias dependent, depicting to the fact that the Fe-Fe, Fe-Ni, Fe-Co ordering can not be asserted as universal, over the entire bias range. On the other hand, numerical calculations, probing their ordering at the representative bias of V=1, showed that the calculated photon-assisted current enhancement is 22.6% for Fe–Co, 10.4% for Fe–Ni, while Fe–Fe shows a 4.4% reduction (Table 5).

Table 5. Comparison between static current, driven current, and enhancement for Fe-Fe, Fe-Ni, and Fe-Co.

Static current (Fe-Fe)	1.677374e-05 A
Driven current (Fe-Fe)	1.602785e-05 A
Relative current change(Fe-Fe)	-0.044467 (-4.447 %)
Static current (Fe-Ni)	1.923191e-06 A
Driven current (Fe-Ni)	2.123273e-06 A
Relative current change(Fe-Ni)	0.104037 (+10.404 %)
Static current (Fe-Co)	1.330410e-06 A
Driven current (Fe-Co)	1.631574e-06 A
Relative current change (Fe-Co)	0.226369 (+22.637 %)

Thus, obtained results reproduces the manuscript's originally stated qualitative ordering at that particular bias (V = 1V), although it is bias-dependent, and given hierarchical ordering is not preserved for the all biases. At the representative bias V, the calculated enhancement follows Fe–Co > Fe–Ni > Fe–Fe. However, the full bias-dependent response demonstrates that this ordering is not universal. At V = 1.0 V, Fe–Co exhibits the largest enhancement among the three systems.

Fe–Ni exhibits moderate enhancement due to intrinsic asymmetry but weaker exchange effects.

Fe–Fe, despite strong magnetism, shows the smallest current amplification because of symmetric Floquet splitting and reduced orbital selectivity.

Finally, quasi-energy spectra at k=0, represented in Fig. 3, 6, and 9, enabled discussion about resonance structure of compared dimers. The quasi-energy spectra at k = 0 provided a compact visualization of resonance behavior. Based on Fig. 3, 6, and 9. Fig. 3, describing 1D Fe–Fe case, showed symmetric level splitting dominated by exchange effects. Fig. 6 (Fe–Ni case), exhibited uneven level spacing reflecting intrinsic onsite asymmetry. Fig. 9 (commenting Fe–Co case) displayed the largest avoided crossings, indicating strongest photon-mediated coupling.

The magnitude of level repulsion provides a qualitative indication of Floquet mixing strength, which is relevant to photon-assisted transport processes.

Accordingly, the combined Model I and Model II calculations illustrate the transport implications of Floquet mixing, orbital hybridization, exchange splitting, and open-junction transport conditions on the photon-assisted current.

4. CONCLUSION

This study demonstrates that combining Floquet band analysis with open-system transport modeling, provides a framework for examining how magnetic Fe–X motifs respond to periodic driving. The calculations show that orbital hybridization, exchange splitting, and Floquet mixing can strongly influence photon-assisted transport, while the magnitude and even the sign of the current modification depend on the applied bias. At the representative bias V = 1.0 V considered here, the Fe–Co parameterization exhibits the largest current enhancement among the three investigated motifs. The results suggests

that effective photon-assisted transport can benefit from the combination of strong hybridization, spin-dependent Floquet replicas, and appropriate resonant band matching. The unified theoretical framework offers a pathway toward rational design of light-controlled molecular magnetic junctions.

Acknowledgments

The research was funded by the Ministry of Education, Science, and Technological Development of the Republic of Serbia (451-03-136/2025-03/ 200017).

Data and Code Availability

The MATLAB simulation code used for the study is available upon request and will be uploaded to a public repository at a later date.

References

[1] SANVITO, S.: Molecular spintronics, *Chemical Society Reviews*, **40** (2011) 3336.

[2] FARAJPOUR, B. E., JAROS, A., BADRI, Z., TUCKOVA, L., STRAKA, M., FOROUTAN-NEJAD, C.: Spinristor: A Swiss Army Knife of Molecular Electronics, *Journal of American Chemical Society*, **143** (2021) 12345.

[3] DELPRAT, S., GALBIATI, M., TATAY, S., QUINARD, B., BARRAUD, C., PETROFF, F., …, MATTANA, R.: Molecular spintronics: the role of spin-dependent hybridization. *Journal of Physics D: Applied Physics*, **51** (2018) 473001.

[4] GUO, L., HU, S., GU, X., ZHANG, R., WANG, K., YAN, W., SUN, X.: Emerging spintronic materials and functionalities. *Advanced Materials*, **36** (2024) 2301854.

[5] ZENG, T. S., SHENG, D. N.: Prethermal time crystals in a one-dimensional periodically driven Floquet system, *Physical Review B*, **96** (2017) 094202.

[6] TUNG, J. C., GUO, G. Y.: Systematic ab initio study of the magnetic and electronic properties of all 3d transition metal linear and zigzag nanowires, Physical Review B, 76 (2007) 094413.

[7] JALILIAN J., ZAHRABI H., JALILIAN J., SOOFIVAND, F., FARSHADFAR, S., NADERIZADEH, S., RAHIMI, N.: Electronic and magnetic properties of Fe, Co and Ni atomic chains encapsulated in BN nanotube bundle, Computational and Theoretical Chemistry, 979 (2012) 10-16.

[8] MATHWORKS, INC.: MATLAB, version R2024a, The MathWorks Inc., Natick, Massachusetts, USA, 2024.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

THERMAL BEHAVIOR OF THE HIERARCHICAL STRUCTURE OF MORPHO DIDIUS BUTTERFLY WINGS: IMPLICATIONS FOR PASSIVE INFRARED SIGNATURE MANAGEMENT

MARINA SIMOVIĆ - PAVLOVIĆ
Teleoptic Gyroscopes, Filipa Višnjića 31, 11080 Belgrade, Serbia, simovicmarina99@gmail.com
Faculty of Mechanical Engineering, University in Belgrade, Kraljice Marije 16, 11000, Belgrade, Serbia
Nanophotonics Lab, Photonics Center, Institute of Physics, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

LUKA MARJANOVIĆ
Teleoptic Gyroscopes, Filipa Višnjića 31, 11080 Belgrade, Serbia, marjanovic.luka98@gmail.com

KATARINA NESTOROVIĆ
Military Technical Institute, Ratka Resanovića 1, 11030, Belgrade, Serbia, katarina.nestorovic993@gmail.com
Military Academy, University of Defense, Veljka Lukića Kurjaka, 11042, Belgrade, Serbia

DARKO VASILJEVIĆ
Nanophotonics Lab, Photonics Center, Institute of Physics, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia, darkovas@gmail.com

Abstract: *Modern surveillance and detection systems increasingly rely on thermal imaging cameras that detect objects based on temperature differences between the target and the surrounding environment. Military vehicles are particularly vulnerable to this type of detection due to significant heat generation from engines, mechanical friction, and exhaust systems, as well as heat transfer to external structural surfaces. As a result, these vehicles often exhibit a pronounced infrared (IR) signature, making them easily detectable under low-visibility or nighttime conditions.*

Recent research has therefore focused on natural structures exhibiting distinctive thermal behavior that may provide principles for passive infrared signature management. In this study, the hierarchical structure and thermal behavior of Morpho didius butterfly wings were investigated using scanning electron microscopy, digital holographic interferometry, differential scanning calorimetry, and infrared thermal imaging. The results revealed pronounced structural and thermal heterogeneity, including temperature-dependent heat capacity, non-uniform surface temperature distribution, and localized thermal inertia under external heating. These findings indicate that the hierarchical organization of the wing strongly influences heat storage, transfer, and spatial temperature distribution.

The identified structure–property–response relationship provides design principles for the development of bioinspired materials and surfaces capable of controlling heat distribution and delaying thermal equilibration. Such architectures may contribute to reduced infrared contrast under transient thermal conditions, providing a potential basis for passive thermal management and infrared signature reduction in military and other defense platforms.

Keywords: *IR signature reduction, thermal imaging detection, thermal masking, military vehicle protection, biomimetic material.*

1. INTRODUCTION

The continuous advancement of thermal imaging technology has greatly improved the performance of modern surveillance, reconnaissance, and target acquisition systems. In contrast to conventional optical sensors, thermal cameras detect infrared radiation emitted by objects according to their surface temperature. This capability allows reliable operation at night, in poor visibility, and under adverse weather conditions. As a result, thermal imaging has become one of the most effective methods for detecting, identifying, and tracking military vehicles on the modern battlefield [1, 2].

As thermal imaging sensors continue to improve in sensitivity and detection range, reducing the infrared (IR) signature of military vehicles has become increasingly important. [3] A vehicle's IR signature is primarily generated by heat sources such as the engine, exhaust system, tracks, and heated external surfaces, which can be detected by modern thermal sensors even under low-visibility conditions. [4] Effective IR signature reduction decreases the probability of detection, identification, and

Corresponding author: M. Simović - Pavlović;
E-mail: simovicmarina99@gmail.com
DOI: 10.5937/OTEH26105S

tracking by enemy surveillance systems and infrared-guided weapons, thereby improving vehicle survivability on the battlefield.

A wide range of techniques has been proposed to mitigate the infrared signature of military platforms, including thermal barrier materials, low-emissivity coatings, exhaust suppression systems, active thermal management, and adaptive camouflage technologies. Although these approaches can substantially decrease the detectability of a target in the infrared spectrum, many of them introduce additional challenges related to system weight, power demand, operational complexity, and maintenance. As a result, recent research has increasingly focused on passive infrared signature management solutions based on advanced materials and engineered surface structures. [5]

Nature offers numerous examples of highly effective thermal regulation mechanisms that have evolved to help organisms survive in diverse and often extreme environments. Many biological systems possess sophisticated hierarchical structures that enable precise control of heat absorption, insulation, dissipation, and redistribution across multiple length scales. Inspired by these naturally optimized solutions, researchers have increasingly explored bioinspired design strategies for the development of advanced materials with tailored thermal properties. [6, 7] Such approaches provide valuable opportunities for creating passive thermal management and infrared camouflage materials capable of regulating surface temperature and heat transfer without external energy input.

In this work, several biological structures known for their distinctive thermal characteristics were examined to explore mechanisms that could be applied to infrared signature management. Experimental thermal analyses were conducted to evaluate their ability to influence heat accumulation, transfer, and spatial temperature distribution. By identifying the underlying principles responsible for these thermal responses, the study aims to contribute to the development of bioinspired materials and surface designs for passive thermal control. Such solutions may offer promising opportunities for reducing the thermal detectability of military vehicles and improving their survivability against modern infrared surveillance and targeting systems. [8]

2. MATERIALS AND METHODS

2.1. Biological Sample

The biological sample investigated in this study consisted of wings of the butterfly *Morpho didius*. The specimens were stored under ambient laboratory conditions in sealed glass containers, without any additional preparation prior to the experiments. The analyses were performed directly on the preserved wings.

Morpho didius was selected because of its highly ordered hierarchical wing architecture, which has been extensively recognized for its unique optical and structural properties. [9] The wing surface is covered with overlapping scales composed of complex micro- and nanoscale features, including multilayered ridge-lamella structures and air gaps. Beyond their well-known role in structural coloration, these nanostructures are expected to influence heat transfer, heat storage, and the spatial distribution of temperature across the wing surface. Their combination of geometry and material organization makes *Morpho didius* an attractive biological model for investigating passive thermal regulation mechanisms relevant to bioinspired infrared signature management.

2.2 Morphological Characterization

Standard microscopic techniques were employed as supporting methods for the characterization of the wing surface morphology (Figure 1). The microstructural features of the *Morpho didius* wings were analyzed using a scanning electron microscope (SEM), JEOL JSM-6610 LV (Japan). SEM imaging was performed in order to resolve the hierarchical micro- and nanoscale architecture of the wing surface. These observations provide a structural basis for interpreting the thermal behavior of the wing in subsequent analyses.

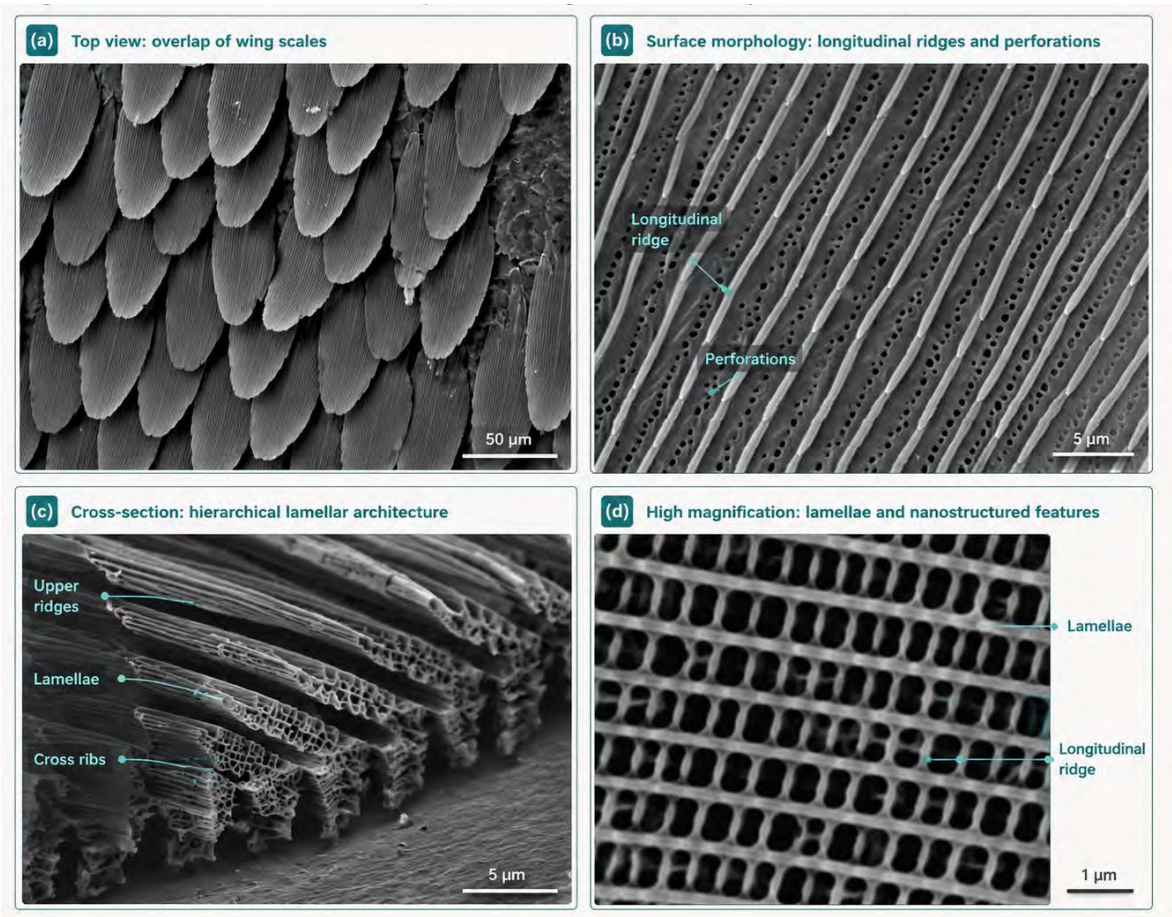


Figure 1. Scanning electron microscopy (SEM) characterization of *Morpho didius* wing scales at multiple hierarchical levels.

(a) Top-view of overlapping cover scales illustrating their roof-like arrangement.
(b) Surface morphology of an individual scale showing longitudinal ridges and perforated architecture.
(c) Cross-sectional view revealing the multilayered ridge–lamella organization responsible for the hierarchical microstructure.
(d) High-magnification image of the lamellar architecture, highlighting the periodic structural features that contribute to the wing's optical and thermal behavior.

In addition to SEM analysis, digital holographic interferometry was employed as a complementary technique for structural characterization of the *Morpho didius* wing. [9] Holography enables the recording of object deformation between two states of exposure, providing full-field, three-dimensional information about structural changes. In this study, it was used to monitor the dynamic response of the wing microstructure under external thermal stimulation. This approach allows observation of nanoscale displacements induced by heating in real time, providing insight into the mechanical and structural behavior of the hierarchical wing architecture. A simplified holographic setup was used for the measurements, while the general configuration of the system has been described elsewhere [10, 11].

2.3 Thermal Characterization

Differential scanning calorimetry (DSC) analysis was performed to determine the heat capacity of *Morpho didius* wing samples. Measurements were carried out using a DSC Q1000 instrument (TA Instruments, Delaware, USA). The samples were analyzed over a temperature range of approximately 22-140 °C, following the standard procedure provided by the manufacturer and in accordance with ASTM E1269-11. The obtained data were used to evaluate the temperature-dependent thermal response of the biological material. Particular features of the heat-capacity curve were observed in the low-temperature range of approximately 22-30 °C and in the range of approximately 86-95 °C.

Infrared thermal measurements were performed in order to observe the surface temperature response of the *Morpho didius* wing under controlled external heating. A FLIR A65sc thermal camera was positioned in front of the sample to record temperature distribution during laser irradiation. The sample was illuminated using a low-power 532 nm laser source, providing controlled localized heating of the wing surface. Thermal images were recorded with a spatial resolution of 640 × 512 pixels and a frame rate of 30 Hz. The data were processed using FLIR ResearchIR Max software (v4.40), enabling analysis of temperature distribution and temporal evolution of the thermal response.

3. RESULTS AND DISCUSSION

3.1 Morphological structure of *Morpho didius* wing

The microstructural organization of *Morpho didius* wing samples was analyzed using scanning electron microscopy (SEM), revealing a highly ordered hierarchical architecture composed of overlapping scales with well-defined micro- and nanoscale features. The surface is characterized by periodic ridge-lamella structures forming a multilayered arrangement with embedded air gaps, which together generate a complex three-dimensional morphology. Such structural organization results in a highly anisotropic surface, where geometry varies significantly across different length scales.

The observed nanoscale features are arranged in a repeating pattern that contributes to the overall structural stability and functional behavior of the wing. In particular, the combination of ridge spacing, lamella thickness, and internal voids suggests a heterogeneous medium in which local structural variations may influence both mechanical and thermal transport processes. This hierarchical organization is a key characteristic of biological photonic structures and is not limited to optical functionality alone.

In addition to static morphological analysis, digital holographic interferometry was used as a complementary technique to observe the dynamic response of the wing structure under external thermal stimulation. The results indicate that localized regions of the wing exhibit measurable nanoscale displacements when subjected to heating, suggesting that the structure is not mechanically rigid at all scales but rather exhibits subtle thermally induced deformation behavior. This dynamic response further confirms the complexity of the hierarchical architecture and its sensitivity to external energy input.

Overall, the combination of SEM and holographic observations demonstrates that the *Morpho didius* wing possesses a multi-scale structural organization with both static and dynamic characteristics. Such hierarchical complexity is expected to play an important role in governing heat distribution and energy transport within the material, providing a structural basis for the thermal behavior analyzed in the following sections.

3.2 Heat capacity behavior of *Morpho didius* wing

The thermal response of *Morpho didius* wing samples was investigated using differential scanning calorimetry (DSC), and the resulting heat capacity (Cp) as a function of temperature is presented in Figure 2. The obtained curve exhibits a pronounced non-linear behavior over the investigated temperature range (20–140 °C), indicating a temperature-dependent thermal response of the biological material.

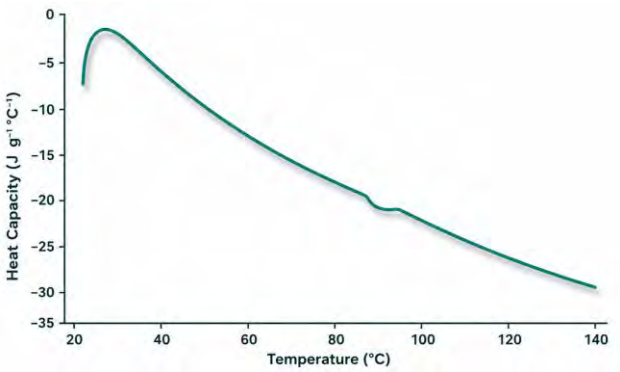


Figure 2. Heat capacity as a function of temperature for the *Morpho didius* wing obtained by differential scanning calorimetry (DSC). The observed nonlinear trend reflects the temperature-dependent thermal behavior of the hierarchical wing architecture and provides the basis for interpreting the subsequent infrared thermal response.

In the low-temperature region (approximately 22–30 °C), a pronounced change in heat capacity is observed, followed by a gradual decrease over the subsequent temperature range up to approximately 85 °C. This behavior may be associated with the relaxation of hierarchical micro- and nanoscale components within the wing architecture.

At approximately 86-95 °C, a distinct change in the slope of the curve is observed, followed by a continued gradual decrease in heat capacity toward 140 °C. This feature indicates a localized change in the thermal response of the material within this temperature range.

Overall, the non-linear temperature dependence of heat capacity indicates that the thermal response of the wing is not uniform over the investigated temperature range. The distinct changes observed in the approximately 22–30 °C and 86–95 °C regions suggest that different temperature intervals are associated with different thermal response characteristics of the biological material.

From an application perspective, such behavior is of particular interest for passive thermal regulation concepts, as temperature-dependent heat capacity may contribute to delayed heat propagation and reduced thermal contrast under transient heating conditions. These features are relevant for the design of bioinspired materials intended for infrared signature management.

3.3 Infrared thermal response

The infrared thermal response of the *Morpho didius* wing was investigated under controlled laser heating conditions using a FLIR A65sc thermal imaging camera. The recorded thermographic data provide a spatial and temporal representation of surface temperature distribution during external energy input, allowing analysis of heat propagation across the wing structure.

Upon laser irradiation, the wing exhibits a non-uniform temperature distribution, indicating heterogeneous heat absorption and dissipation across the surface. Localized regions respond differently to thermal excitation, resulting in spatial temperature gradients that reflect the underlying

hierarchical micro- and nanoscale architecture. The observed thermal behavior suggests that heat transfer within the structure is not homogeneous, but rather governed by localized structural variations.

Time-resolved thermal imaging further reveals that the cooling dynamics of the wing are non-uniform, with certain regions retaining heat for longer periods than others. This indicates the presence of localized thermal inertia, likely induced by structural complexity and air-trapped micro- and nanoscale features. Such behavior is consistent with the presence of multi-scale architectures that influence both heat storage and dissipation pathways.

The combination of spatial and temporal thermal variations observed in the FLIR measurements supports the existence of structure-dependent thermal regulation mechanisms within the wing. These findings are in agreement with the DSC results, which indicate temperature-dependent heat capacity behavior, suggesting that both macroscopic thermal response and thermodynamic properties are governed by the same hierarchical structural organization.

Overall, the infrared thermal response demonstrates that the *Morpho didius* wing does not behave as a thermally uniform material, but rather as a complex system with spatially varying heat transfer characteristics. This behavior is relevant for understanding passive thermal regulation in biological systems and may provide inspiration for the development of bioinspired materials for infrared signature management.

To integrate the morphological features, thermal properties, and observed infrared response into a unified interpretation, a schematic overview of the system behavior is presented in Figure 3.

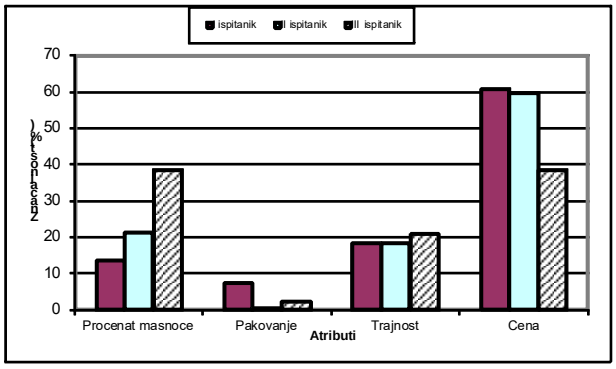


Figure 3. Structure-property-response framework illustrating how the hierarchical architecture of the *Morpho didius* wing governs its temperature-dependent thermal behavior and influences the apparent infrared response, providing inspiration for passive infrared signature management concepts.

4. ENGINEERING IMPLICATION

The observed structure–property–response relationship suggests design principles for bioinspired materials capable of advanced thermal regulation (Figure 4).

The observed thermal behavior may be particularly relevant to military platforms exposed to transient heating conditions, where infrared detectability is strongly influenced by the temperature distribution of external surfaces. The bioinspired approach proposed in this study could be translated into engineered multilayer or hierarchical surface structures designed to redistribute and temporarily store heat, thereby reducing local temperature gradients and delaying thermal equilibration. Such structures could potentially be incorporated into passive thermal-management layers, surface treatments, or camouflage-related materials applied to selected external components of military vehicles and other defense platforms. The present study does not experimentally demonstrate infrared signature reduction on a military vehicle; rather, it provides a biological model and experimentally observed structure–property–response relationships that may serve as a basis for the future design and testing of such materials under realistic military operating conditions.

Hierarchical architectures similar to those found in *Morpho didius* wings may enable a decoupling between bulk thermal properties and apparent infrared response, where the observed thermal signature is governed primarily by structural organization rather than homogeneous material behavior. Such mechanisms can lead to controlled heat storage and dissipation, resulting in reduced thermal contrast under transient heating conditions. This indicates that engineered surfaces based on similar principles could exhibit thermal behavior that is not directly representative of the overall system temperature, but instead reflects localized structural interactions with heat transfer processes. From an application perspective, such behavior is relevant for infrared signature management and the development of passive systems aimed at reducing thermal detectability of complex platforms.

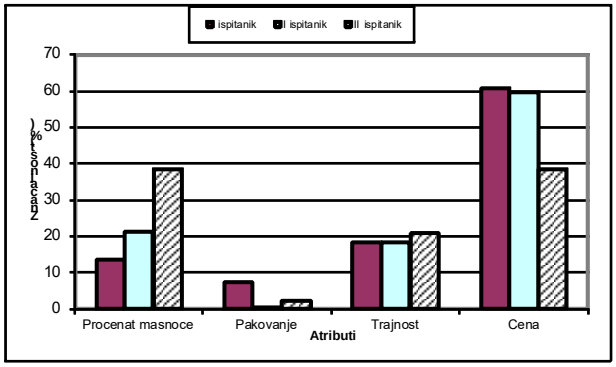


Figure 4. Bioinspired concept illustrating the relationship between hierarchical structural organization, heat redistribution, and apparent infrared response. The proposed mechanism demonstrates how structural design may govern thermal emission independently of bulk material properties, providing a framework for passive infrared signature management.

5. CONCLUSION

This study investigated the hierarchical structure and thermal behavior of *Morpho didius* butterfly wings to

identify structural principles relevant to passive thermal regulation and infrared signature management. The combined morphological and thermal analyses revealed a pronounced structure-dependent thermal response, characterized by temperature-dependent heat capacity, non-uniform surface temperature distribution, and localized thermal inertia under external heating.

The most significant finding is that the thermal response cannot be described solely by homogeneous bulk material properties. Instead, the hierarchical organization of the wing strongly influences heat storage, transfer, and spatial temperature distribution, resulting in a structure-dependent apparent infrared response. This structure–property–response relationship represents the principal functional insight obtained from the investigated biological system.

These findings provide a basis for bioinspired thermal management concepts in which hierarchical surface architectures are designed to control heat redistribution and delay thermal equilibration. Such an approach may offer a potential route toward reducing infrared contrast under transient thermal conditions and developing passive materials for infrared signature management.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] STRICKLAND, R.N.: I*nfrared Techniques for Military Applications*, Infrared methodology and technology, CRC Press, 2023. (397-427)

[2] GERKEN, M.C., MÜNZBERG, M.O.: *Novel trends in optronic imaging*, Infrared Sensors, Devices and Applications XV, SPIE, 13613 (2025).

[3] BARANWAL, N., MAHULIKAR, S.P.: *Review of infrared signature suppression systems using optical blocking method,* Defence Technology, 15(3), (2019), 432-439.

[4] PLESA, C., TURCANU, D., BODOC, V.: *The use of infrared radiation for thermal signatures determination of ground targets,* Romanian Journal of Physics, 51(1/2), (2006), 63.

[5] SIMOVIC PAVLOVIC, M., PAGNACCO, M., BOKIC, B., VASILJEVIC, D., RADMILOVIC-RADJENOVIC, M., RADJENOVIC, B., KOLARIC, B.: *Breaking barriers: Molding thermodynamics by geometry of nanostructures,* In 2024 24th International Conference on Transparent Optical Networks (ICTON) (pp. 1-4). IEEE.

[6] SIMOVIC PAVLOVIC, M., BOKIC, B., VASILJEVIC, D., KOLARIC, B.: *Bioinspired NEMS – Prospective of collaboration with nature,* Applied Sciences, 12(2), (2022), 905.

[7] LING, X., OSOTSI, M.I., ZHANG, W., WU, Y., JIN, Q., ZHANG, D.: *Bioinspired materials: From distinct dimensional architecture to thermal regulation properties*, Journal of Bionic Engineering, 20(3), (2023), 873-899.

[8] KARIIS, H., SAVAGE, S.: *Signature management – Trends, materials and concepts 2017,* FOI-R-4509-SE, Swedish Defence Research Agency (FOI), 2018.

[9] SIMOVIC PAVLOVIC, M., GRUJIC, D., PAGNACCO, M., BOKIC, B., VASILJEVIC, VERBIEST, T., KOLARIC, B.: *Thermophoretic effect in natural photonics: holographic study,* Bioinspiration & Biomimetics, 20(2), (2025), 024001.

[10] PANTELIC, D.V., GRUJIC, D.Z., VASILJEVIC, D.M.: *Single-beam, dual-view digital holographic interferometry for biomechanical strain measurements of biological objects,* Journal of Biomedical Optics, 19(12), (2014), 127005-127005.

[11] SIMOVIC PAVLOVIC, M.C., PAGNACCO, M., GRUJIC, D., BOKIC, B., VASILJEVIC, MOUCHET, S., VERBIEST, T., KOLARIC, B.: *Uncovering hidden dynamics of natural photonic structures using Holographic imaging,* JoVE (Journal of Visual Experiments), 181, (2022), e63676.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EXPERIMENTAL EVALUATION OF THE TENSILE PROPERTIES OF DIFFERENT TYPES OF ARMOUR STEEL AND THEIR WELDED JOINTS

VLADIMIR MILOVANOVIĆ
Faculty of Engineering, University in Kragujevac, Sestre Janjić 6, 34000, Kragujevac, Serbia, vladicka@kg.ac.rs

DUŠAN ARSIĆ*
Faculty of Engineering, University in Kragujevac, Sestre Janjić 6, 34000, Kragujevac, Serbia, dusan.arsic@fink.rs

DUŠAN POPADIĆ
Zastava Tervo, Kosovska 4, 34000, Kragujevac, Serbia, dule.pop@gmail.com

MILOŠ PEŠIĆ
Institute for Information Technologies, University of Kragujevac, Jovana Cvijića bb, Kragujevac, Serbia, milospesic@uni.kg.ac.rs

KATARINA DRAGUTINOVIĆ
Faculty of Engineering, University in Kragujevac, Sestre Janjić 6, 34000, Kragujevac, Serbia, katarinaristicina@gmail.com

Abstract: *This paper presents the results of an experimental investigation into the mechanical properties and quality assessment of welded joints and their corresponding base materials. The study focuses on high-strength armour steels, a class of steel that due to their high hardness and toughness under dynamic impact loads, are utilized in specialized protective structures. The primary objective was to evaluate the performance of the welded joints produced by a specific welding technology and MAG welding process and to compare their characteristics with those of the unaffected base metal. Tensile testing was conducted on a universal testing machine equipped with appropriate instrumentation for precise measurement of force and elongation. A total of 27 specimens were tested, including both transverse welded samples and base metal reference samples, ensuring a comprehensive dataset. The tests allowed to determine key mechanical parameters, including yield strength, ultimate tensile strength, percentage elongation after fracture, and reduction of area. Beyond the numerical results, a fractographic analysis was performed, with particular attention given to the location and morphology of the fracture. The fracture surfaces were examined to determine whether failure occurred in the weld metal, the heat-affected zone (HAZ), or the base material, providing critical insights into the weakest link of the joint. The obtained results were analysed to assess the efficiency of the welding MAG procedure.*

Keywords: *welded joints, tensile testing, mechanical properties, armoured high-strength steels, fracture analysis, heat-affected zone (HAZ)*

1. INTRODUCTION

The modern development of Infantry Fighting Vehicles (IFVs) stemmed from the continuous effort to enhance tank operational efficiency and battlefield survivability. This evolutionary requirement presented a major metallurgical and design challenge: developing armour systems capable of defeating small-arms fire and anti-tank projectiles while simultaneously maintaining optimal tactical, technical, and operational performance. Meeting these stringent criteria necessitated the development of specialized ultra-high-strength steels, commonly referred to as armour or ballistic steels, which continue to undergo active technological refinement [1-3]. Prominent among these materials is the commercial ARMOX series produced by the Swedish manufacturer SSAB Oxelösund, manufactured under strictly controlled production protocols [4]. Their primary function is to defeat kinetic energy threats such as small arms fire, armour-piercing projectiles, and explosive shrapnel while maintaining minimum structural weight to preserve vehicle mobility and operational efficiency. The superior ballistic and mechanical characteristics of ARMOX steels result directly from their tailored chemistry and manufacturing processes: ultra-low carbon content (C<0.3%) formulated specifically to improve field weldability and minimize cold-cracking susceptibility. Thermomechanical processing (TMP): utilized to achieve exceptionally high yield and ultimate tensile strengths [5,6]. However, despite these superior properties, joining armour steel plates into a coherent hull structure introduce localized material degradation. The intense heat input during the welding process inevitably alters thermo-mechanically processed microstructure, creating a localized softening zone or microstructural mismatch in the heat-affected zone (HAZ). Consequently, these welded joints represent potential critical structural vulnerabilities when exposed to dynamic or ballistic loading conditions. To address this

Corresponding author: V. Milovanović;
E-mail: vladicka@kg.ac.rs
DOI: 10.5937/OTEH26106M

challenge, this study presents a comprehensive experimental investigation into the mechanical properties, fracture behaviour, and overall quality assessment of ARMOX welded joints compared directly to the base metal.

2. MATERIALS AND METHODS

The objective of this research is to determine fundamental mechanical parameters (Re, Rm, percentage elongation A, and reduction of area Z) for different welded samples using specific welding technology. Based on obtained results, researchers investigated how welded joints perform under tensile loads, assess the technological efficiency of the applied welding procedure, and provide key insights into the structural integrity of armoured vehicles when subjected to severe mechanical stress. Testing was done by using a universal testing machine equipped with precise force and elongation instrumentation. A total of 27 specimens, comprising both transverse welded samples and base metal reference standards were tested to destruction. For testing and welding two different armour (ballistic) steels were used as well as one structural steel:

1. ARMOX 500T,
2. SWEBOR 500,
3. S355J0.

Besides armour steels, S355J0 steel was tested because in common military use the armour steels are often welded directly to structural steel, so the authors wanted to test the properties of that welded joints too.

In tables 1 and 2 the chemical composition and mechanical properties are given.

Table 1. Chemical composition of ARMOX 500T, Swebor 500 and S355J0

Steel type	Chemical composition, %									
	C	Si	Mn	Cr	Mo	P	S	Mo	Ni	B
ARMOX 500T	0.32	0.5	1.3	1.0	0.7	0.01	0.005	0.7	1.8	0.005
Swebor 500	0.30	0.5	0.4	NS	NS	0.015	0.005	NS	/	/
S355J0	0.22	0.55	1.6	/	/	0.035	0.035	/	/	/

*NS – not specified.

Table 2. Mechanical properties of G42CrMo4 cast steel [8]

Steel	R_m, MPa	$R_{p0.2}$, MPa	A, %	Z, %	Impact energy, J
ARMOX 500T	1650	1300	7-10	/	25 (R)
Swebor 500	1670	1200	9.5	/	25 (-40°C)
S355J0	510-680	355	14-18	/	27 (-20°C)

After materials selection, we approached to samples preparation. Samples were prepared in company Zastava TERVO, Kragujevac, Serbia.

For the purposes of testing and analysing the obtained results, base material samples of the selected steels were prepared, as well as test specimens for welded joints with all possible combinations. For initial testing purposes, a total of twenty-seven samples were prepared (shown in Fig. 1), according to standard SRPS EN ISO 6892-1 [7] and drawing given in Fig. 2, consisting of 9 different series with 3 samples in each series:

- Series 1 - Base material samples of S355 structural steel – 3 pcs.
- Series 2 - Base material samples of Swebor 500 high-strength steel – 3 pcs.
- Series 3 - Base material samples of Armox 500 high-strength steel – 3 pcs.
- Series 4 - Welded samples of S355 structural steel – 3 pcs.
- Series 5 - Welded samples of S355 structural steel and Swebor 500 high-strength steel – 3 pcs.
- Series 6 - Welded samples of S355 structural steel and Armox 500 high-strength steel – 3 pcs.
- Series 7 - Welded samples of Swebor 500 high-strength steel – 3 pcs.
- Series 8 - Welded samples of Armox 500 high-strength steel – 3 pcs.
- Series 9 - Welded samples of Swebor 500 and Armox 500 high-strength steels – 3 pcs.

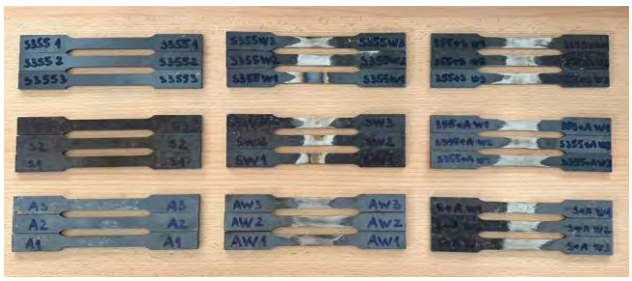


Figure 1. Prepared samples

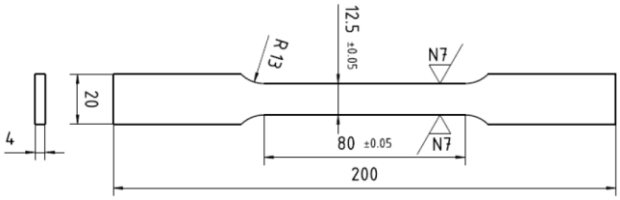


Figure 2. Drawing for samples preparation

Before testing the specimens to determine the mechanical properties of dissimilar steel welded joints, the mechanical properties of the base metals of the selected steels were determined.

The specimens for determining the mechanical properties

of dissimilar steel welded joints for all tested series were prepared through a sequence of manufacturing operations. The test specimen parts were first prepared for the welding process. Following laser cutting, the edges were bevelled so that the specimen parts could be welded. Each specimen was welded using the same method and with identical welding parameters.

Welding was performed with a runoff tab (backing tab) to prevent the formation of defects during the strike and extinction of the electric arc over such a short distance. After welding, the specimen was ground smooth at the joint location (Fig. 3). The welding was carried out using the MAG process with MIGWELD 307Si austenitic filler wire (diameter 1.0 mm) and M12 shielding gas (Ar + 0.5% CO_2). This filler wire was selected to prevent brittle fracture in the weld metal, as recommended by both the manufacturer [4] and other authors [6]. The chemical composition and mechanical properties of the filler wire are given in Table 3.

Figure 3. Appearance of welded sample

Table 3. Chemical composition of filler metal MIGWELD 307Si

Chemical composition, %				
C	Si	Mn	Cr	Ni
0.08	<1.0	7.0	18.5	9.0
Mechanical properties				
Rm, MPa	Rp0,2, MPa	A, %	KV, J	
560-660	>380	>35	32 (-196°C)	

For tensile testing, the Shimadzu servo-hydraulic testing machine (pulsator) (Fig. 4) and the MFA25 extensometer (MF Mess & Feinwerktechnik GmbH, Velbert, Germany) with a gauge length of 50 mm (Fig. 5) were used. Direct integration with the Servo 4830 controller and software enables automated data acquisition and real-time graphical visualization throughout each test. Following specimen mounting, the extensometer was positioned at the midpoint of the gauge length to measure real deformation. The device was carefully attached to prevent slippage during testing..

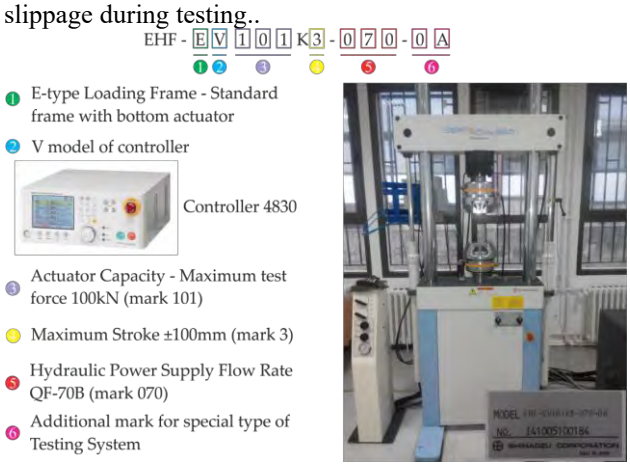


Figure 4. SHIMADZU servo-hydraulic machine

Figure 5. MFA25 extensometer

Tensile testing of the specimens was conducted at room temperature (23±5°C) at a constant displacement rate of 0.9 mm/min (Fig. 6). Prior to each test, zeroing of all measurement channels on both the testing machine and within the Servo 4830 software was carried out to ensure the accuracy and reliability of the recorded data.

Figure 6. Sample during testing

3. RESULTS AND DISCUSSION

Based on the tests performed on the specimens, an analysis of the obtained results was conducted. Following data processing using the DataViewer 4830 software, engineering stress–displacement curves were generated for each uniaxial tensile test. Furthermore, engineering stress–engineering strain curves along the A50 extensometer gauge length can also be plotted, as extensometer opening—and thus strain—was recorded continuously throughout all experiments.

For result analysis and comparison, engineering stress–displacement curves are presented for each tested specimen (Fig. 7). This approach was adopted because fracturing and necking did not consistently occur within

the extensometer gauge length. As can be clearly seen from the figure, the specimens fabricated from the S355 structural steel baseline material exhibit the highest elongation. Conversely, the high-strength steels, Swebor 500 and Armox 500, demonstrate high yield strength and tensile strength values, accompanied by lower elongation compared to the S355 structural steel.

To facilitate tracking and evaluation of the obtained mechanical properties across all tested specimens, the complete set of results is summarized in Table 3.

Analysis of the obtained results revealed that the mechanical properties of the welded joints significantly depend on the base metal type. For the S355 structural steel, relatively uniform plastic deformation was observed, with fracture occurring primarily in the base metal or in the immediate vicinity of the heat-affected zone (HAZ). Conversely, due to their higher hardness and specific microstructure, the armour steels exhibited reduced plastic deformation prior to failure, alongside a more pronounced localization of stress and fracture within the weld zone.

The test results indicate that the proper selection of welding parameters and filler material plays a vital role in determining the quality and load-bearing capacity of the welded joint, particularly for high-strength steels. Furthermore, it was confirmed that the heat-affected zone (HAZ) represents a critical region of the welded structure, as it undergoes substantial microstructural alterations and degradation of mechanical properties.

The conclusions made by this analysis is mostly aligned with the conclusions of other authors which find that superior ballistic and mechanical properties of ARMOX steels to a fully tempered martensitic microstructure, which simultaneously imposes inherent limitations regarding ductility and strain hardening capacity [1]. Under tensile and dynamic loading, these materials are particularly prone to stress localization and premature failure driven by stress triaxiality and local plastic instability [3]. When joining such ultra-high-strength steels, thermal cycles during welding inevitably induce structural heterogeneity across the joint zones, leading to localized variations in mechanical resistance. To mitigate the risk of brittle failure within the weld metal—a critical drawback highlighted in high-hardness steel joining—austenitic or ductile filler metals are routinely implemented [6]. Consequently, the observed reduction in overall elongation and the stress concentration within specific joint regions for example HAZ/weld metal, align with literature findings, confirming that the soft austenitic weld matrix absorbs plastic strain while protecting the joint against catastrophic brittle fracture.

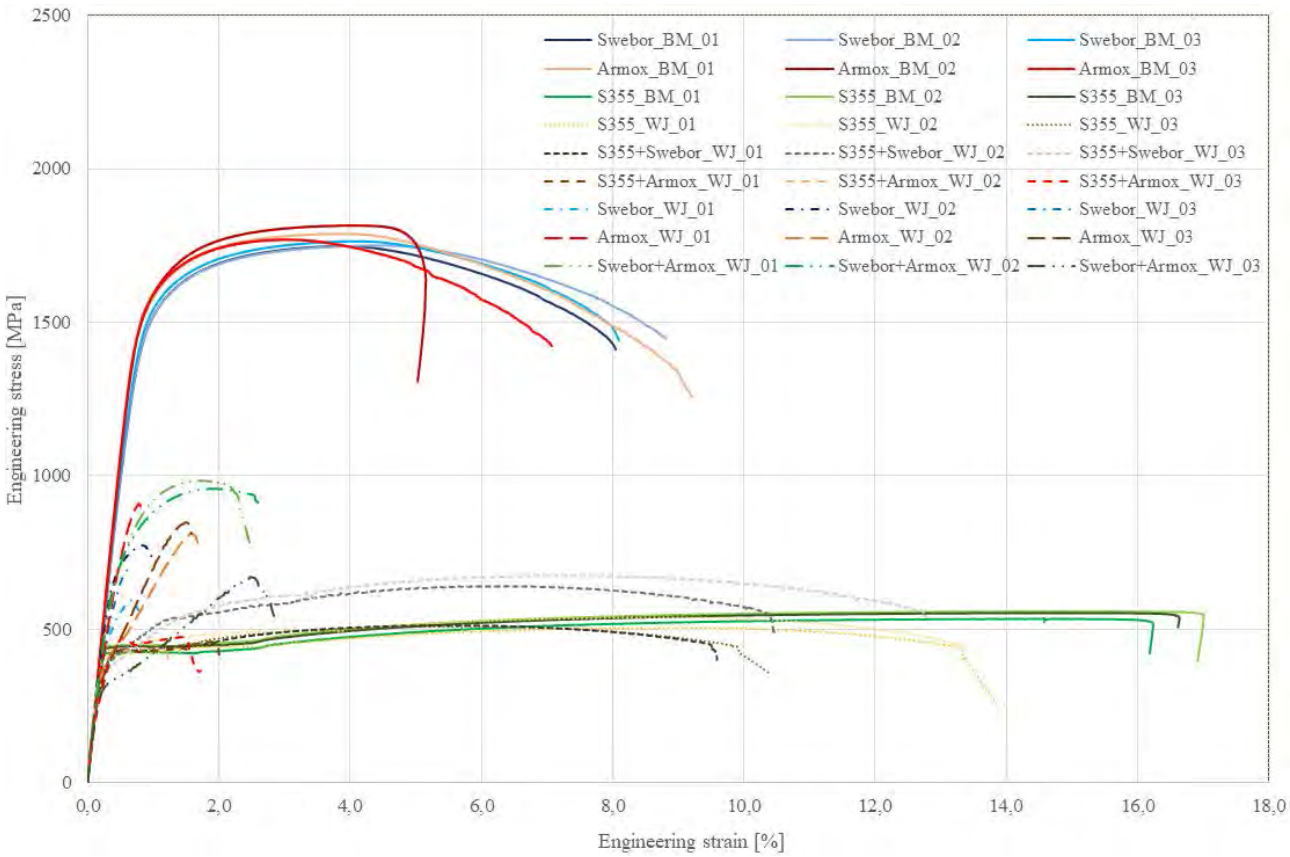


Figure 7. Stress-strain diagrams for tested specimens

Table 3. Results obtained by tensile testing

Sample no.	Elasticity modulus E, MPa	Yield stress, $R_{p0,2}$, MPa	Tensile strength, R_m, MPa	Fracture location
Swebor_OM_01	204273	1484.4	1748.7	-
Swebor_OM_02	203203	1487.0	1751.0	-
Swebor_OM_03	204054	1509.6	1764.6	-
Armox_OM_1	217619	1541.4	1788.8	-
Armox_OM_2	215319	1555.2	1816.2	-
Armox_OM_3	221284	1553.0	1770	-
S355_OM_1	176623	422.08	535.28	-
S355_OM_2	182713	447.02	558.13	-
S355_OM_3	180622	441.79	553.05	-
S355_ZS_1	182192	415.33	504.01	HAZ
S355_ZS_2	181723	427.21	536.98	HAZ
S355_ZS_3	186641	428.8	510.41	HAZ
S355+Swebor_1	182963	441.03	511.05	Welded joint
S355+Swebor_2	200073	359.26	414.01	Welded joint
S355+Swebor_3	191195	429.07	511.19	HAZ
S355+Swebor_4	207138	401.37	640.53	HAZ
S355+Swebor_5	219892	379.93	673.67	HAZ
S355+Armox_1	199375	439.74	459.86	Welded joint
S355+Armox_2	192213	426.11	442.8	Welded joint
S355+Armox_3	183336	445.41	473.65	Welded joint
Swebor_ZS_01	192244	547.68	594.15	Welded joint
Swebor_ZS_02	204824	724.21	772.16	Welded joint
Swebor_ZS_03	190049	677.57	693.26	Welded joint
Armox_ZS_01	196462	828.23	910.5	Welded joint
Armox_ZS_02	189000	776.57	813.56	Welded joint
Armox_ZS_03	190255	665.86	781.45	Welded joint
Armox_ZS_04	198142	435.81	809.08	Welded joint
Armox_ZS_05	206256	426.02	850.07	Welded joint
Swebor+Armox_01	194510	764.65	983.91	Welded joint
Swebor+Armox_02	190867	547.41	562.89	Welded joint
Swebor+Armox_03	186946	764.32	958.21	Welded joint
Swebor+Armox_04	179842	331.85	669.12	Welded joint
Swebor+Armox_05	180418	375.16	741.96	Welded joint

5. CONCLUSION

Based on the experimental investigations and the analysis of the results, several key conclusions can be drawn regarding the mechanical behaviour of dissimilar welded joints under tensile loading. The tests were performed on welded joints made of S355 structural steel and two types of armour steels (Armox 500 and Swebor 500), aiming to evaluate the influence of the base metal on the load-bearing capacity, plastic deformation capability, and fracture behaviour of the welded joint.

The experimental results demonstrated that the mechanical properties of the welded joints significantly depend on the mechanical and metallurgical characteristics of the base metal. For the S355 structural steel, a relatively uniform stress distribution was observed during tension, along with a pronounced plastic deformation capacity prior to fracture (Fig. 7). In the majority of tested cases, fracture occurred in the base metal or in the immediate vicinity of the heat-affected zone (HAZ), indicating that the weld metal quality was not the critical weakness of the structure.

The results for the armour steels revealed a specific mechanical behaviour resulting from their high hardness and strength. These materials exhibited reduced plastic deformation capacity alongside pronounced stress localization within the heat-affected zone (HAZ), specifically in the coarse-grained HAZ adjacent to the fusion line. Such behaviour can lead to earlier fracture initiation, particularly in regions subject to microstructural alterations induced by the welding heat input.

A critical aspect of the analysis concerns the heat-affected zone (HAZ), where significant microstructural and mechanical changes take place. The findings confirm that the HAZ represents one of the most critical regions in welded structures, particularly for steels with high mechanical properties.

Consequently, it can be concluded that welded joints fabricated from S355 steel display favourable mechanical behaviour and good ductility, making them suitable for structural components subjected to static loads. Conversely, armour steels—despite their high strength and resistance to mechanical damage—require stringent control of welding parameters due to their heightened sensitivity to microstructural changes.

Test specimens comprised both similar joints (same base metals) and dissimilar joints (combinations of the aforementioned steels). Particular attention was devoted

to the preparation of the weld zone to ensure the representativeness of the results and to minimize the influence of geometric imperfections.

Acknowledgement

The authors are thankful to the company Zastava Tervo, Kragujevac for providing the necessary resources for samples preparation.

References

[1] KHAN, W., TUFAIL, M.; CHANDIO, A.D. *Characterization of microstructure, phase composition, and mechanical behavior of ballistic steels*, Materials, 15 (2022), 2204. https://doi.org/10.3390/ma15062204

[2] LAZIĆ, V., ARSIĆ, D., RAKIČEVIĆ, M., NIKOLIĆ, R., UHRIČIK, M., HADZIMA, B., *Influence of the welding joint type on safety properties of the armor steel ARMOX 500T*, System Safety: Human - Technical Facility – Environment, 1(1) (2019) 753-759. https://doi.org/10.2478/czoto-2019-0096

[3] POPŁAWSKI, A., KĘDZIERSKI, P., MORKA, A., Identification *of Armox 500T steel failure properties in the modelling of perforation problems*, Materials & Design, 190 (2020) 108536, https://doi.org/10.1016/j.matdes.2020.108536.

[4] https://www.ssab.com/en/brands-and-products/armox/product-offer/armox-500t

[5] POPADIĆ, D. Experimental determination of mechanical properties of different welded joints of various steels, master thesis, University of Kragujevac, Faculty of Engineering, Kragujevac, Serbia, 2026.

[6] GARASIC, I., REMENAR, M., BARIŠIĆ, A., ILJKIĆ, D., Determination of ballistic properties on ARMOX 500T steel welded joint, Engineering review, 39(2) (2019) 186-196, DOI: 10.30765/er.39.2.8

[7] SRPS EN ISO 6892-1:2020, Metallic materials - Tensile testing - Part 1: Method of test at room temperature (ISO 6892-1:2019).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

FLEXIBLE ELECTROCHEMICAL ENERGY SOURCES FOR WEARABLE SYSTEMS: TECHNICAL SOLUTIONS AND INTEGRATION WITH A MILITARY APPLICATION ASPECT

BOŽANA GAĆIĆ
Technical Test Center, Belgrade, boska93va@gmail.com

MIHAEL BUČKO
University of Defense Belgrade, Military Academy, mihael.bucko@va.mod.gov.rs

MILICA KOŠEVIĆ
University of Belgrade, Institute of Chemistry, Technology and Metallurgy - National Institute of the Republic of Serbia, milica.kosevic@ihtm.bg.ac.rs

RADOVAN KARKALIĆ
University of Defense Belgrade, Military Academy, radovan.karkalic@va.mod.gov.rs

SRETEN PERIĆ
University of Defense Belgrade, Military Academy, sreten.peric@va.mod.gov.rs

Abstract: *The development of wearable electronic systems has led to an increased need for compact, lightweight and mechanically flexible energy sources that can be integrated into textile materials. This paper analyzes flexible electrochemical energy sources intended for wearable systems, with a special focus on technical solutions, materials and possibilities for their integration into smart textiles. The main categories of flexible energy devices are discussed, including flexible batteries, supercapacitors and fiber energy systems, as well as their advantages over traditional batteries. Special attention is given to the role of advanced nanomaterials, such as graphene and MXene, which allow improvement of the electrochemical and mechanical properties of these devices. Different methods of integrating energy components into textiles are also analyzed, including weaving conductive fibers, printing functional materials and coating fibers. The paper also discusses the potential of these systems in military wearable technologies, where it is necessary to provide reliable power supply for numerous electronic devices with minimal increase in the mass of the equipment. The analysis presented in this paper indicates that flexible electrochemical energy sources represent a promising solution for the development of a new generation of smart textiles and wearable systems.*

Keywords: *flexible batteries, supercapacitors, wearable electronics, MXenes, textile integration.*

1. INTRODUCTION

The modern development of wearable electronic systems has led to a significant increase in the demand for reliable, lightweight, and energy-efficient power sources. In military applications, wearable electronic systems include communication devices, navigation systems, biomedical sensors, and systems for real-time data acquisition and processing. Traditional batteries exhibit limitations in the context of wearable systems, primarily due to their rigid structure, limited mechanical flexibility, and relatively high mass. These drawbacks affect mobility, ergonomics, and operational efficiency in military conditions [1].

In this context, flexible electrochemical energy sources represent a promising solution that enables the combination of electrochemical performance and mechanical adaptability [2]. The development of these systems is based on the application of advanced materials, such as nanostructured composites and two-dimensional materials, as well as innovative structural approaches that enable integration into textile structures. The integration of energy systems into smart textiles is of particular importance, as it enables uniform mass distribution, improved comfort, and the application of self-powered devices [3].

The aim of this paper is to provide a comprehensive analysis of flexible electrochemical energy sources, primarily flexible batteries and supercapacitors, with an emphasis on their performance and design constraints. Special attention is given to their comparison with conventional rigid battery systems, particularly in terms of mechanical flexibility and operational stability under deformation.

Furthermore, the paper examines recent technical solutions enabling flexibility, including novel architectures and the use of advanced materials such as MXene and other conductive nanostructures. A dedicated focus is placed on strategies for integrating these energy

Corresponding author: B. Gaćić;
E-mail: boska93va@gmail.com
DOI: 10.5937/OTEH26107G

sources into textile platforms, including weaving conductive fibers, printing functional materials, and coating fibers with electrochemically active layers.

Finally, the study explores the applicability of these systems in wearable technologies, with particular emphasis on military applications.

2. FLEXIBLE ENERGY SOURCES

Flexible electrochemical energy storage devices may be defined as systems typically composed of electrodes, current collectors, electrolytes, separators, and supporting substrates, designed to storage chemical energy and preserve performance under bending, twisting, folding, and stretching [4]. The rapid development of portable and wearable electronics, including wireless sensors, flexible displays, electronic skins, and foldable devices, has created a growing demand for these systems.

To enable fully autonomous wearable devices, flexible energy storage technologies must provide high energy density, long lifespan, fast charging and discharging capability, and operational safety at an acceptable cost [5]. In addition, these systems must maintain stable performance under mechanical deformation, requiring both electrical reliability and mechanical durability. As shown in Figure 1, significant research efforts have been directed toward designing materials and architectures capable of withstanding bending, twisting, folding, and stretching without compromising functionality [1].

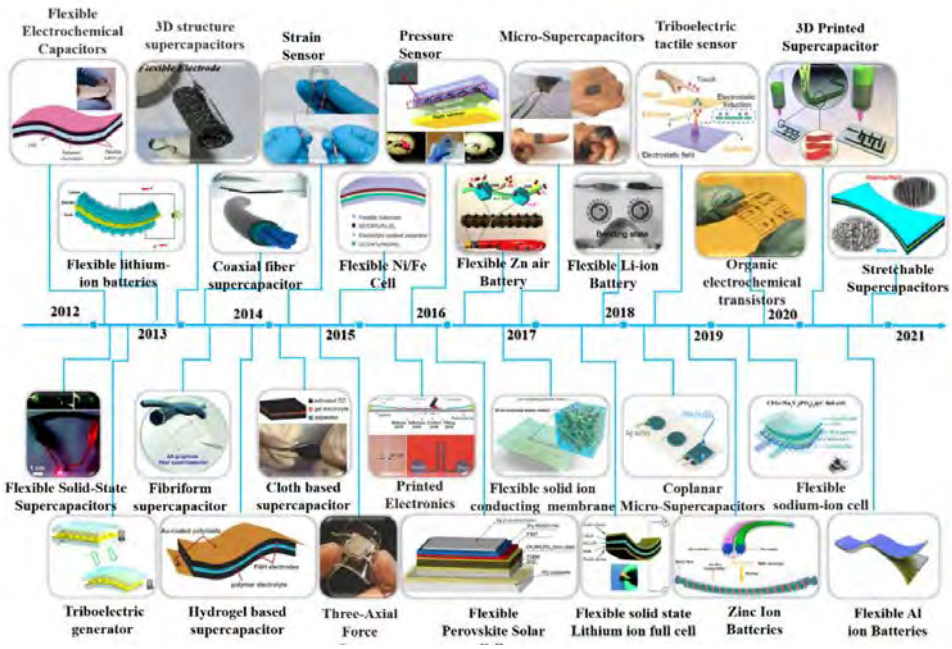


Figure 1. The evolution of flexible energy storage devices [1]

Flexible electrodes play a key role in meeting these requirements and are often designed without binders, conductive additives, or rigid current collectors. Namely, conventional electrode materials are typically rigid and mechanically brittle, limiting their use in flexible devices. Additionally, inactive binders may reduce capacity and introduce safety concerns, leading to the development of binder-free electrode designs [1].

Self-charging systems combining flexible energy storage with energy harvesting devices have also been developed, allowing wearable systems to achieve partial or continuous energy autonomy during operation [5]. The application of such storage + harvesting concept is illustrated in Figure 2 [2].

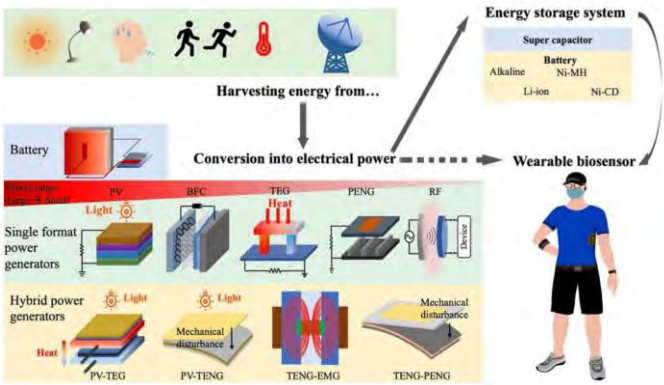


Figure 2. Schematic flow diagram of self-powered smart wearable sensors. Harvesting energy from various energy sources, converting into electric power, storing energy in storage section and powering wearable sensors. PV: photovoltaic; BFCs: biofuel cells; TEGs: thermoelectric generators; PENGs: piezoelectric nanogenerators; TENGs: triboelectric nanogenerators; RF: radio frequency; EMGs: electromagnetic generators [2]

Figure 2 illustrates the pathway from harvesting energy from various sources to its conversion into electrical energy, storage, and use for powering wearable sensors. The depicted energy sources include light, heat, radio waves, and vibrations. Photovoltaic (PV) systems are used for harvesting light energy, while biofuel cells (BFCs) represent an electrochemical energy source. Thermoelectric generators (TEGs) are used for harvesting thermal energy, whereas piezoelectric nanogenerators (PENGs) and triboelectric nanogenerators (TENGs) are technologies for harvesting mechanical energy. Radio-frequency (RF) systems utilize radio-frequency energy sources, while electromagnetic generators (EMGs) represent an additional technology for energy harvesting. The harvested energy is converted into electrical power, stored, and subsequently used to supply power to wearable sensors [2].

3. COMPARISON WITH TRADITIONAL BATTERIES

Traditional batteries, particularly lithium-ion batteries, remain the dominant energy storage technology due to their relatively high energy density and proven stability in portable electronic systems [1,4]. However, their rigid structure and limited mechanical flexibility represent a significant limitation for applications in wearable systems [1,2]. Flexible electrochemical energy sources have been developed to overcome these limitations.

While conventional batteries retain an advantage in energy density, flexible supercapacitors and hybrid systems exhibit higher power density and rapid charge/discharge capability, making them suitable for short-term but intensive energy demands in wearable devices [5,6]. This indicates that flexible energy sources are not a direct replacement for conventional batteries, but rather their complement in specific wearable applications.

Additionally, flexible systems enable new device architectures, such as fiber-based and textile-based configurations, which are not possible with traditional batteries. These approaches allow better integration into wearable systems, more uniform mass distribution, and improved ergonomics [7,8]. Furthermore, as highlighted in [9], distributed textile-based battery systems represent a paradigm shift compared to centralized conventional batteries. Instead of relying on a single rigid energy source, energy can be spatially distributed across clothing, significantly improving redundancy, fault tolerance, and ergonomic performance.

4. TECHNICAL SOLUTIONS FOR FLEXIBLE ELECTROCHEMICAL ENERGY SOURCES

The two most important flexible energy storage systems have been rechargeable batteries (flexible alkaline, zinc–carbon, polymer, etc.) and supercapacitors. Flexible rechargeable batteries offer high energy density but are limited by relatively low power density and slower charge/discharge rates. In contrast, supercapacitors provide high power density, rapid charge/discharge capability, and long cycle life, making them suitable for complementary applications.

4.1 Devices

4.1.1 Flexible Batteries

Flexible batteries represent one of the most important solutions for powering wearable electronic systems. The modern development of these systems is based on modifying conventional lithium-ion batteries using thin-film structures, polymer electrolytes, and flexible current collectors [4].

Two dominant structural configurations of flexible energy sources are ultrathin flexible structures (UFS) and one-dimensional fiber-based architectures (Figure 3). Depending on fiber density and structural orientation, fiber-based batteries can be classified as coaxial, multi-axial, or ultrathin flexible structures. Each configuration offers different advantages in terms of mechanical flexibility and electrochemical performance [4].

As shown in Figure 3, the following configurations can be observed: (a) coaxial one-dimensional structure, (b) multi-axial one-dimensional structures, (c) ultrathin flexible structure for AMAD, (d) spine-like structure, (e) interwoven structure, (f) node-type structure for AMAD, (g) serpentine interconnection, (h) origami and kirigami structures, and (i) coplanar electrode structure for IAMA [6].

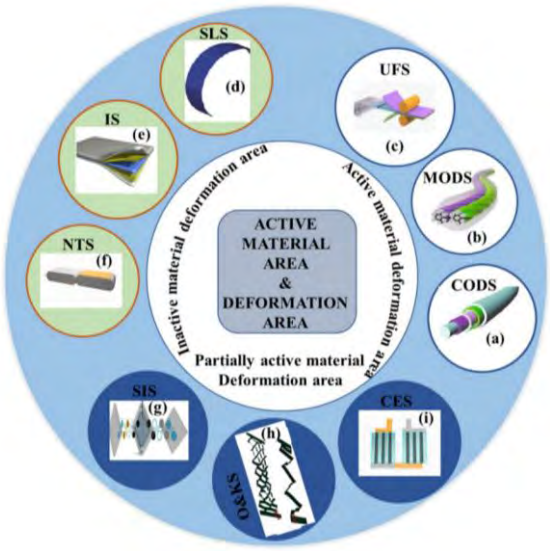


Figure 3. Structural designs for flexible and stretchable batteries, including active and inactive regions and deformation zones [4]

The integration of textile-based triboelectric nanogenerators (TENGs) with flexible batteries enables the development of self-charging hybrid systems. These systems can convert biomechanical energy generated during human motion into electrical energy, which is then stored in flexible batteries, extending operational time without external charging infrastructure [10].

Advanced design approaches include the development of fiber-based and textile-integrated batteries, as well as printed battery systems that can be applied in flexible wearable devices [9]. Additionally, techniques such as screen printing and inkjet printing allow deposition of functional materials onto textile substrates [7,8].

4.1.2 Supercapacitors

Flexible supercapacitors represent an alternative to batteries in applications requiring high power density and rapid charging, making them suitable for energy-demanding wearable systems [6]. Significant progress has been achieved in the development of fiber-shaped supercapacitors that can be directly integrated into textile structures. These systems are typically realized in coaxial, twisted, or layered configurations [11].

A schematic overview of supercapacitor operation, the number of publications in this field, and their integration into wearable electronics are shown in Figure 4 [6].

Fabric-based flexible supercapacitors show strong potential for integration into multifunctional wearable systems due to their high-power density and mechanical durability. Such systems can be directly embedded into military clothing, enabling energy delivery to components such as communication modules and sensor arrays [12].

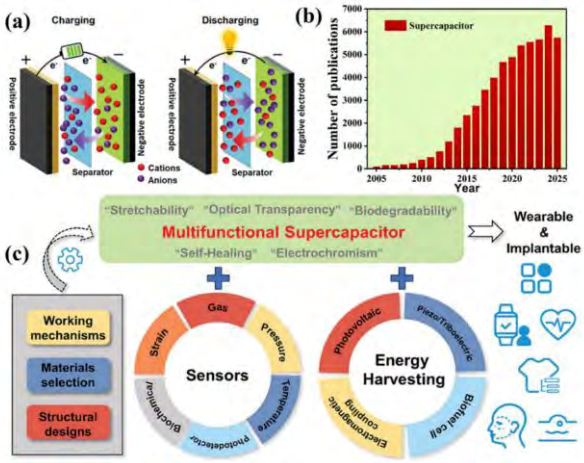


Figure 4. Overview of supercapacitor integration with sensors and electronic devices: (a) energy storage mechanisms, (b) publication trends, (c) schematic integration with wearable and implantable systems [6]

4.2 Advanced Materials

4.2.1 MXenes

Two-dimensional transition metal carbides, nitrides, and carbonitrides known as MXenes have emerged as a revolutionary class of materials for the development of multifunctional wearable electronic textiles [13]. Due to their unique structure and tunable surface functional groups, MXenes exhibit a range of exceptional properties, as shown in Figure 5 [13]. Their outstanding mechanical flexibility, hydrophilicity, tunable surface chemistry, and high electrical conductivity make them ideal for integration into textile systems [13].

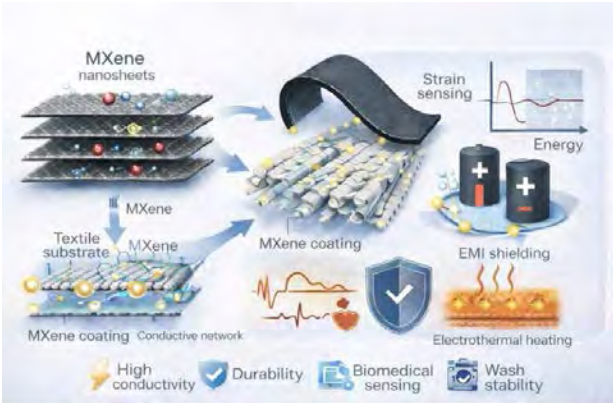


Figure 5. Structure and properties of MXenes [13].

Thanks to their large specific surface area and excellent electrical conductivity, MXenes enable improved energy and power density in flexible energy devices [14]. Particular attention has been given to their application in electrochemical energy storage, especially in supercapacitors. MXenes are highly suitable electrode materials due to their superior mechanical flexibility, energy density, and electrochemical performance [15].

Six main categories of MXene–polymer nanocomposites, classified based on dispersion state and macroscopic form, are shown in Figure 6 [16].

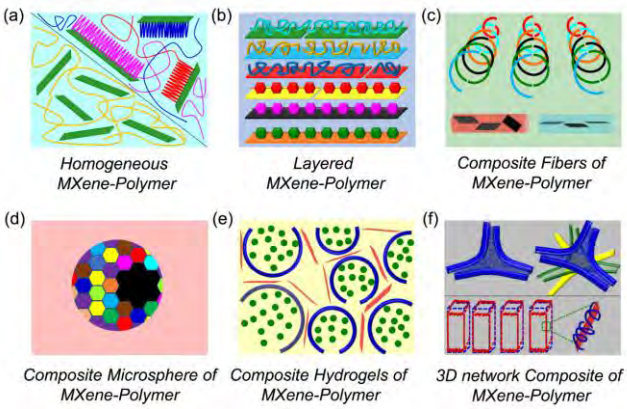


Figure 6. The Six Main Types of MXene–Polymer Nanocomposites [16]

4.2.2 Graphene

Graphene, a two-dimensional material composed of a single layer of sp^2-hybridized carbon atoms, has emerged as a key advanced material for flexible electrochemical energy storage systems. Owing to its exceptional electrical conductivity, high specific surface area, and outstanding mechanical strength, graphene exhibits a unique combination of properties that makes it highly suitable for wearable electronic applications. In addition, its intrinsic flexibility and low density enable seamless integration into textile-based platforms, supporting the development of lightweight and mechanically adaptable energy storage devices [7].

Due to its high specific surface area and excellent charge transport properties, graphene is widely used as electrode

material in flexible supercapacitors and batteries, enabling high power density through electric double-layer capacitance. It also maintains stable performance under mechanical deformation, which is essential for wearable systems. However, challenges such as sheet restacking and low intrinsic pseudocapacitance can limit its overall energy storage performance. To address these limitations, graphene is often combined with nanostructured materials and conductive polymers, forming composite systems with improved ion transport, enhanced capacitance, and better mechanical stability [17].

The schematic illustration (Figure 7) presents typical graphene-based composite electrode architectures, including polymer-functionalized and hybrid nanostructures. These designs enhance ion transport and prevent restacking, resulting in improved electrochemical performance and flexibility, which are essential for wearable and textile-integrated energy storage systems [17].

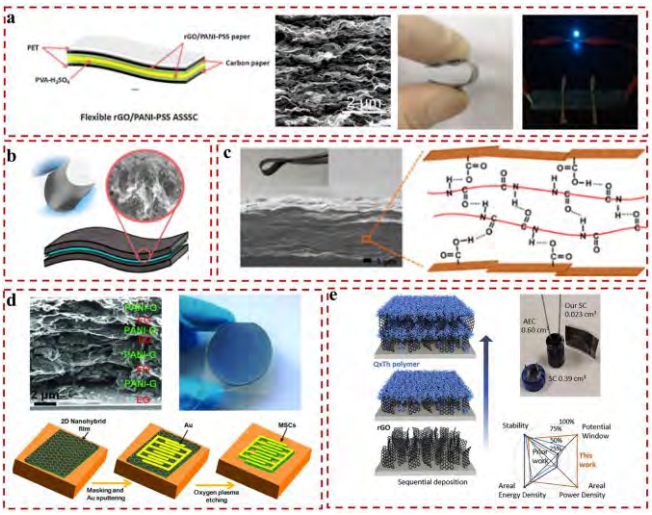


Figure 7. Flexible graphene-based supercapacitors [17]

5. INTEGRATION INTO TEXTILES

The integration of flexible electrochemical energy sources into textile structures represents a key step in the development of wearable systems, enabling direct incorporation of energy functions into clothing and equipment. This approach provides improved ergonomics, reduced weight, and uniform distribution of functional components, which is particularly important in military applications [8]. Modern integration strategies include weaving conductive fibers, printing functional materials, and coating textile substrates, often combining multiple methods to achieve optimal performance [7].

Printed distributed energy storage systems enable seamless integration into textile architectures without significantly altering mechanical properties, which is crucial for maintaining comfort and flexibility in long-term military use. [9]

5.1 Weaving Conductive Fibers

Weaving conductive fibers is one of the most direct approaches for integrating functional components into textiles.

Fiber-based supercapacitors and conductive fibers based on graphene, MXene, and conductive polymers can be incorporated during weaving or knitting processes, forming functional fabrics capable of storing and delivering energy [8].

5.2 Printing Functional Materials

Printing techniques provide a flexible and precise approach for integrating electrochemical components into textiles. Methods such as inkjet printing, screen printing, and direct ink writing (DIW) enable direct deposition of functional materials onto textile substrates, allowing fabrication of electrodes, interconnections, and energy devices with high geometric control [8,18]. Printing enables the development of thin-film, lightweight and flexible systems that can be adapted to complex textile structures.

Inkjet printing involves controlled deposition of microdroplets containing conductive or electrochemically active materials, enabling thin and uniform layers with high precision [18].

Screen printing allows thicker and more uniform layers, offering advantages in simplicity, cost-effectiveness, and scalability [18].

Direct ink writing (DIW) enables extrusion of viscous inks into continuous or 3D structures, allowing precise control over electrode architecture and mechanical stability [18].

5.3 Coating of Fibers with Functional Materials

Coating represents one of the simplest methods for integrating functional materials into textiles. Techniques such as dip-coating, spray-coating, and vacuum filtration enable deposition of thin conductive and electrochemically active layers while preserving textile flexibility [7,13].

Dip-coating involves immersion of the substrate in a solution, forming a uniform conductive layer after drying. Complex multilayer systems such as PU@PMA@EGaIn exhibit high conductivity (>10^5 S/m) and stretchability up to 500% [7].

Spray coating enables controlled deposition with efficient material usage, producing CNT–graphene/PU fibers with conductivity up to 8 S/m and stretchability up to 620% [7].

Vacuum filtration produces uniform nanocomposite films, such as aramid nanofiber/Ag nanowire papers with low surface resistance (0.12 Ω/sq) and high thermal stability (>500 °C), which represent model systems applicable also in flexible textile structures [7].

Figure 8. MXene/Ni coating process on PET fabrics [13]

By combining these methods, it is possible to achieve advanced functional structures, as illustrated in Figure 8. For example, the formation of a bilayer MXene/Ni coating on polyester fabric resulted in a significant improvement in electrical conductivity (up to 113.8 S/cm), hydrophobicity, and electromagnetic interference shielding effectiveness, while preserving the porosity of the material [13].

6. MILITARY APPLICATION

After discussing the integration of flexible energy devices into textile structures, it is important to consider their practical role in military applications. Modern wearable military systems require reliable, lightweight, and efficient energy sources to power communication devices, navigation systems, and biomedical sensors [2]. Flexible electrochemical energy sources provide a more uniform distribution of mass across uniforms and equipment, while supporting advanced wearable concepts such as smart uniforms with integrated systems for monitoring health status and tactical parameters [5].

In practical military applications, these technologies involve the direct integration of energy systems into uniforms, vests, and soldier equipment, allowing multiple devices to be powered without reliance on a centralized energy source [9].

These approaches are aligned with NATO soldier modernization programs, such as the “Future Soldier System” and related initiatives, which aim to digitalize and network the individual soldier as a combat platform. Within these systems, the soldier operates as a carrier of multiple electronic subsystems (communications, positioning, sensors), significantly increasing energy demands and necessitating distributed and flexible power solutions.

A specific application example includes powering wearable sensors for monitoring physiological parameters such as body temperature, heart rate, and fatigue level. These data enable real-time assessment of a soldier’s combat readiness, representing a key component of modern soldier system concepts [12].

Furthermore, NATO initiatives emphasize reducing logistical burden, where flexible and hybrid energy systems can significantly decrease the number of spare batteries a soldier must carry. The integration of multiple energy sources (storage + harvesting) enhances reliability, redundancy, and continuity of power supply under diverse operational conditions [12].

Overall, flexible electrochemical energy sources represent a key enabling technology for future soldier systems, as they support direct integration of energy functionalities into equipment, increase energy autonomy, and reduce logistical constraints.

7. CONCLUSION

Flexible electrochemical energy sources represent a key technological foundation for next-generation wearable electronic systems, particularly in smart textiles and military applications.

Their primary advantage lies in combining electrochemical performance with mechanical adaptability, enabling reliable operation under real-world conditions.

Comparative analysis shows that flexible systems complement rather than replace conventional batteries, forming hybrid energy architectures combining high energy density (batteries) and high power density (supercapacitors).

Advanced materials, particularly MXenes and their nanocomposites, significantly enhance conductivity, capacity, and mechanical stability.

Integration strategies such as fiber-based systems, printing, and coating enable distributed energy systems directly within textiles.

In military applications, these technologies support the “soldier-as-a-system” concept by reducing weight, improving ergonomics, and increasing energy autonomy.

Future research should focus on robust hybrid systems, interface optimization, and scalable manufacturing technologies.

Acknowledgment

The authors would like to acknowledge support from the scientific project of the Military Academy, University of Defense in Serbia, “Research on influence of characteristics of explosive ordnance on safety in Ministry of Defense and Army of Serbia” (VA-TT/1/22-24), and from the Ministry of Science, Technological Development and Innovation of Republic of Serbia (Contract No. 451-03-33/2026-03/200026).

References

[1] ZHANG, Q., GAO, X., LIU, X., MU, J., GU, Q., LIU, Z., LUO, W.: Flexible wearable energy storage devices: Materials, structures, and applications, *Battery Energy*, 3(2) (2024).

[2] RONG, G., ZHENG, Y., SAWAN, M.: Energy Solutions for Wearable Sensors: A Review, *Sensors*, 21(11) (2021).

[3] BI, Y., YU, M., LI, J., WANG, A., XU, H., ZHANG, X.: Review on Self-Sustainable Power Generation Technologies for Future Typical Wearable Applications, *The International Journal of Multiphysics*, 17(3) (2023).

[4] ALI, A., CAI, R., WANG, T., ZHANG, C., LIU, C.,

MA, D., HOU, H., LEI, Q., XIA, Y., WANG, S., WEI, W., TIAN, L.: Emerging flexible battery technology: Innovative structural design, material integration and wearable applications and beyond, *Chemical Engineering Journal*, 519 (2025).

[5] MANOJKUMAR, K., MUTHURAMALINGAM, M., SUNDAMOORTHY, A., SRINIVASA BABU, P. S., VAMSI, R., BOOMINATHAN, J., SRI PRIYANKA, S. G., HAJRA, S., PANDA, S., KIM, H. J., DUBAL, D., VIVEKANANTHAN, V.: Revolutionizing energy storage: Self-charging supercapacitors toward self-powered micro and nano systems, *Chemical Engineering Journal*, 524 (2025).

[6] RANI, S., KUMAR, P., TANDON, A., KUMAR, S., PRODROMAKIS, T.: Recent advancements in supercapacitor-integrated self-powered sensors, *Chemical Engineering Journal*, 530 (2026).

[7] ZHU, C., WU, J., YAN, J., LIU, X.: Advanced Fiber Materials for Wearable Electronics, *Advanced Fiber Materials*, 5(1) (2023) 12–35.

[8] CHEN, J., HE, T., DU, Z., LEE, C.: Review of textile-based wearable electronics: From the structure of the multi-level hierarchy textiles, *Nano Energy*, 117 (2023).

[9] ALI, A. E., JEOTI, V., STOJANOVIĆ, G. M.: Fabric based printed-distributed battery for wearable e-textiles: A review, *Science and Technology of Advanced Materials*, 22(1) (2021) 772–793.

[10] BATRA, R., POURJAFARIAN, N., CHANG, S., TSAI, M., REVELO, J., KAO, C. H.-L.: texTENG: Fabricating Wearable Textile-Based Triboelectric Nanogenerators (2025).

[11] ZHOU, X., ZHENG, X., DONG, X., WANG, X., JIANG, Y., LI, G., LI, L., YUAN, N., DING, J.: Recent advances in wearable fiber-shaped supercapacitors: Materials, design, and applications, *Energy Materials*, 5(5) (2025).

[12] SINGH, A., SHAH, S. S., DUBEY, A., AHMED, A., UD DIN RATHER, M., RANGANAYAKULU, S. V., SUNDRAMOORTHY, A. K., ARYA, S.: Advancements in wearable energy storage devices via fabric-based flexible supercapacitors, *Journal of Energy Storage*, 109 (2025).

[13] RUPANTY, N. S., GHOSH, J., NOOR, T., ASIF, T. R., AHMED, S., HOWLADER, S., REUKOV, V.: Advances in Wearable Technology: MXene-Based Multifunctional and Biomedical Smart Textiles, *ACS Omega*, 11(1) (2026) 197–228.

[14] SHEN, Y., ZHANG, J., GUAN, F., WANG, J., LI, X., ZHANG, T.: Low-temperature-operable MXene/PEDOT/Graphene fiber-based supercapacitors with ultrahigh energy density, *Chemical Engineering Journal*, 526 (2025).

[15] OTGONBAYAR, Z., YANG, S., KIM, I.-J., OH, W.-C.: Recent advances in 2D MXene and solid state electrolyte for energy storage applications: Comprehensive review, *Chemical Engineering Journal*, 472 (2023).

[16] ZHOU, L., LI, J., XING, C.: MXene–polymer nanocomposites for precise structural and functional applications, *Chemical Engineering Journal*, 506 (2025).

[17] DU, X., LI, Y., WANG, H., ZHANG, L.: Graphene-based flexible supercapacitors for wearable energy storage applications, *Journal of Energy Storage*, 75 (2024)

[18] ALI, S. M., NOGHANIAN, S., KHAN, Z. U., ALZAHRANI, S., ALHARBI, S., ALHARTOMI, M., ALSULAMI, R.: Wearable and Flexible Sensor Devices: Recent Advances in Designs, Fabrication Methods, and Applications, *Sensors*, 25(5) (2025).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

PERFORMANCE AND DEVELOPMENT TRENDS OF PERSONAL BALLISTIC PROTECTIVE EQUIPMENT IN THE CONTEXT OF MODERN THREATS

SINIŠA GAĆIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, gacic.sinisa@gmail.com

NIKOLA PRODANOVIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, nikola.prodan7@gmail.com

MARKO MILOVANOVIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, marko.milovanovic547@gmail.com

Abstract: *Modern conflicts pose new challenges for personal ballistic protection equipment, particularly in the context of urban warfare and the use of unconventional tactical means. This study analyzes the performance of existing protective equipment, trends in the development of new materials and technologies, and the limitations of testing standards. Special attention is given to the use of unmanned aerial vehicles, which increase the risk of fragmentation and airborne threats, placing additional demands on the design, control, and application of protective equipment. Existing standards for ballistic protection testing are critically reviewed, highlighting limitations that may lead to insufficient assessment of effectiveness against current threats. The study also presents an overview of new composite and advanced materials that enable improved protection, lighter and more mobile equipment, as well as the potential application of modular and smart systems. The findings indicate the need for adaptation of standards and improvement of equipment design to ensure higher reliability and user safety under complex tactical and technologically advanced conditions. This study provides a systematic insight into the limitations and potential of modern personal ballistic protective equipment and contributes to a better understanding of its practical application.*

Keywords: *injury mechanisms, tactical requirements, advanced materials, testing standards.*

1. INTRODUCTION

The first quarter of the 21st century has brought radical changes to the nature of armed conflicts, shaped by the technological revolution and new geopolitical realities. Modern battlefields, such as the ongoing conflict in Ukraine, are characterized by asymmetric actions, non-linearity, and the massive use of high-tech systems, blurring the traditional boundaries between the front line and the rear. The dominance of artillery, coupled with the ubiquitous presence of unmanned aerial vehicles and loitering munitions, has presented personal ballistic protective equipment with challenges that exceed previous experiences from conflicts of earlier decades.

The primary issue facing modern protective systems is the discrepancy between advanced threats, materials, and existing testing standards. Injury analyses indicate that while modern vests effectively protect the torso, there is a shift in injury epidemiology toward unprotected body parts, such as the head, neck, and extremities, necessitating changes in equipment design and coverage.

The aim of this paper is a comprehensive performance analysis of existing protective equipment in light of modern threats, with a particular focus on the limitations of current standards. The study explores development trends in new materials, such as nanocomposites and non-Newtonian fluids, which promise enhanced protection with reduced mass and greater user mobility. Through a systematic review of injury mechanisms and technological innovations, the paper seeks to highlight the necessary directions for adapting protective systems to ensure reliability and safety in complex tactical conditions.

2. THREATS IN MODERN WARFARE

The end of the Cold War marked a profound turning point in the way armed forces are prepared, deployed, and equipped. Alongside shifts in military strategy, the defense industry itself underwent a radical transformation, embracing technologies of the Fourth Industrial Revolution to develop and produce advanced weaponry at unprecedented speed, even amid disruptions to global supply chains. Unlike in previous eras, innovations originating in the civilian sector—especially in information and communication technologies, electronics, and advanced materials—are now rapidly transferred to the military domain. This trend has allowed even smaller nations to field highly sophisticated, next-generation weapons systems, narrowing the technological gap on the battlefield.

Corresponding author: S. Gaćić;
E-mail: gacic.sinisa@gmail.com
DOI: 10.5937/OTEH26108G

The evolution of global military technology can be divided into four principal periods, each shaped by distinct security threats, forms of conflict, and dominant weapon systems [1]:

- Bipolar World – Late Cold War, 1970–1991
- Unipolar World – Post-Cold War, 1991–2007
- Omnipolar World – Multicentric, 2008–2022
- Polar Confrontation – "Collective West" vs. "Collective East," 2022–present

The current period is characterized by several notable trends [1]:

- Emergence of cyber threats and intensified non-kinetic operations
- Resurgence of strategic and tactical thermonuclear capabilities
- Emphasis on defensive systems and weapons with indirect effects, surpassing the focus on offensive and direct-effect arms
- Widespread use of unmanned aerial vehicles (UAVs) for both reconnaissance and combat roles
- Expansion of long-range missile systems across all military branches
- Revival of mass conventional ground forces, particularly infantry
- Increased investments in strategic reconnaissance, electronic intelligence, jamming systems, and command-information networks
- Growing focus on hypersonic missile technologies

The current battlefield has also seen the return of mass conventional ground forces, particularly infantry, alongside the growing importance of long-range missile systems, strategic reconnaissance, electronic intelligence, jamming capabilities, and command-information networks at all levels. Hypersonic missile technology has become a priority, reflecting the race to achieve strategic advantages in speed, range, and precision. Artillery has regained primacy on the battlefield, leveraging improved reconnaissance, rapid reaction times, and increasingly sophisticated precision munitions to prevent force concentrations and inflict maximum attrition. Traditional rapid offensive maneuvers have given way to combat resembling fortified or urban warfare, where proximity, complexity, and urban terrain pose heightened operational challenges.

Drones, particularly so-called "killer drones," have emerged as pivotal tools. Their impact stems from mass deployment rather than individual performance, enabling forces to penetrate enemy air defenses, shape battlefield engagements, and achieve operational objectives without risking personnel. Three categories of small drones stand out in terms of lethality: first-person view (FPV) drones, one-way attack (kamikaze) drones, and loitering munitions [2]. These systems effectively function as self-contained missiles. One-way attack drones deliver precise, self-detonating strikes, while loitering munitions hover over target zones, enabling attacks on beyond-line-of-sight objectives. Prominent examples include the U.S. Switchblade 300 and 600, as well as the Russian Lancet and KUB-BLA (Figure 1). Modern drone operations are increasingly treated as consumable assets, deployed not only for destruction but also to disrupt, exhaust, and manipulate enemy forces. Integrating drone intelligence and precision strikes reduces overall ammunition requirements, eases logistical burdens, and accelerates the tempo of artillery operations.

Figure 1. Switchblade 300 (left) and KUB-BLA (right)

The deployment of gliding bombs has also increased. These are traditional aerial bombs modified with deployable wings and guidance systems, allowing for extended range and improved accuracy. Mass production of these weapons reached 40,000 units in 2024 and 70,000 units in 2025 [3], proving decisive in recent conflicts, particularly on the Ukrainian battlefield.

2. DISTRIBUTION AND MECHANISMS OF COMBAT-RELATED INJURIES

In order to evaluate the performance of ballistic protective equipment, comparing injury mechanisms and their anatomical distribution provides critical insights into body armor efficacy. For this purpose, a comparative analysis was conducted between operations from 2001 to 2017 (primarily Iraq and Afghanistan) and the ongoing high-intensity conflict in Ukraine (2022–present).

The earlier period was characterized by asymmetric warfare with a high prevalence of improvised explosive devices (IEDs), standardized personal protective equipment, and near-immediate tactical medical evacuation. In contrast, the operational environment in Ukraine is defined by peer-to-peer conventional warfare, artillery dominance, widespread UAV deployment, variable armor coverage, and protracted casualty evacuation timelines.

While this comparison highlights shifts in dominant injury mechanisms and anatomical distributions driven by evolving threats and protective strategies, certain epidemiological and methodological limitations must be noted. Direct statistical comparison between these eras is confounded by variations in casualty documentation practices, differences in trauma system infrastructure, heterogeneity in worn armor capabilities, and distinct battlefield evacuation dynamics. Consequently, observed shifts reflect not only changes in ballistic armor performance, but also the broader systemic differences of these two operational environments.

2.1 Period from 2001. to 2017.

Analyses of combat injuries sustained during the conflicts in Iraq and Afghanistan indicate that, among 1,992,232 deployed US servicemembers, a total of 29,624 combat wounds were recorded in 7,877 casualties [4]. The anatomical distribution of these injuries consistently shows a predominance of extremity wounds, accounting for 51.9%, followed by injuries to the head and neck region (28.1%), while thoracic and abdominal injuries

each account for around 10%. These findings are corroborated across multiple studies, demonstrating a relatively stable distribution pattern despite variations in operational environments [5, 6].

A defining characteristic of these conflicts is the dominance of explosive mechanisms of injury (MOI), which account for roughly 74.4% of all combat casualties, significantly exceeding the proportion of gunshot wounds (19.9%) [5]. Explosive devices, including improvised explosive devices (IEDs), mortars, rocket-propelled grenades, and landmines, produce complex injury patterns characterized by multiple entry points, widespread tissue damage, and frequent involvement of multiple organ systems. In contrast to gunshot wounds, fragment-related injuries exhibit greater dispersion and are more likely to result in polytrauma.

Within explosive injury mechanisms, extremities remain the most frequently affected anatomical region (41.3%), followed by the head and neck (37.4%), while injuries to the torso are less common (8.8%) [5]. This distribution reflects the protective effect of modern ballistic equipment, which provides substantial coverage of the thorax and abdomen, while leaving the limbs, face, and neck comparatively exposed. Consequently, these unprotected or partially protected regions remain particularly vulnerable to injury, emphasizing persistent limitations in current protective system design. Supporting this observation, postmortem analyses from earlier conflicts indicate that over 20% of fatal penetrating injuries involved the facial region, particularly the midface [7].

Over the course of the Iraq and Afghanistan conflicts, a temporal shift in injury patterns has been observed, particularly concerning head injuries (Tables 1 and 2). The use of more effective protective equipment and body armour is a clear explanation for this shift in anatomical distribution of wounds. Surprisingly, the percentage of extremity injury did not change a lot, while with the protective measures, a decrease might have been expected. The proportion of casualties with head trauma increased significantly in both theaters, rising from approximately 20–30% in early phases of the conflicts to over 50% in later periods [8]. This trend coincides with the increased prevalence of explosive mechanisms and the widespread adoption of advanced body armor systems, which have reduced the incidence of fatal torso injuries but have not proportionally decreased injuries to exposed anatomical regions.

Table 1. Injured body region [8]

Body region	Afganistan			Iraq		
	2001 – 2006	2006 – 2009	2009 – 2017	2003	2004 – 2007	2008 - 2017
Head	30.3	58.8	59.0	20.9	34.8	53.0
Face	25.1	26.0	26.1	25.2	26.3	22.8
Chest	19.2	22.1	19.4	13.0	17.2	19.6
Abdomen	20.3	23.9	24.6	18.2	20.2	16.7
Extremity	51.1	47.4	50.0	47.0	50.9	45.0

Table 2. Mechanisms of injury [8]

Mechanism of injury	Afganistan			Iraq		
	2001 – 2006	2006 – 2009	2009 – 2017	2003	2004 – 2007	2008 - 2017
Explosive	61.4	80.5	77.4	70.1	78.7	79.8
Gunshot	29.1	17.9	21.2	18.4	18.1	16.6
Other	9.5	1.6	1.4	11.5	3.2	3.5

One hallmark of these conflicts has been the increase in explosive devices as primary MOIs. Explosive injuries significantly increase the likelihood of death, with a reported 34.1% increase in odds of killed in action (KIA) death in Afghanistan and a 114% increase in odds of KIA death and 72% increase in odds of DOW (died of wounds) death in Iraq [8]. These mechanisms frequently result in complex, multi-region trauma, further complicating treatment and reducing survival rates.

Traumatic brain injury (TBI) represents a major component of combat-related morbidity in these conflicts. Severe TBIs are most commonly caused by penetrating trauma, including injuries from fragments, shrapnel, and high-velocity projectiles. Approximately 3–6% suffer further infection, and 8–9.5% die due to their injuries [9].

2.2 Period from 2022. to the present

Among Ukrainian soldiers, the most prevalent injuries include extremity, facial, and neck wounds, as well as traumatic brain injury (TBI) [10]. These patterns reflect the high concussive forces associated with modern explosive weaponry, with injury zones predominantly affecting areas not effectively protected by current personal ballistic protective equipment. The conflict demonstrates both an intensification and magnification of the injury patterns previously documented among US personnel during operations in Iraq and Afghanistan, which can be attributed to the types of weapons deployed and the sustained exposure of combatants to high-intensity engagements over prolonged periods.

Modern explosives, including thermobaric devices, have resulted in a broader range and increased severity of traumatic injuries despite protective measures. These weapons are associated with burns, pulmonary trauma, and blast-related TBIs, often circumventing the protective effects of conventional body armour as well as fortifications such as trenches and bunkers. The Russian Federation's use of high-energy explosives has been linked to critical abdominal, vascular, and head injuries,

frequently involving penetrating fragments or secondary blast impacts [10].

A study analyzing 17,692 injured Russian soldiers between 2022 and 2024 on the Ukrainian battlefield evaluated both the tactical delivery platforms and the resulting injury mechanisms across different echelons of care [11]. As presented in Table 3 (derived from a cohort of n=5813 casualties admitted to tertiary military care), three-quarters (75.5 %) of casualties were targeted by UAV systems (including dropped munitions and FPV kamikaze drones), demonstrating the shift in modern firepower delivery. Importantly, UAVs function as tactical delivery platforms rather than distinct biophysical trauma vectors. The explosive payloads delivered by UAVs (such as small mortar shells, shaped charges, and antipersonnel grenades) predominantly inflict high-velocity fragment impacts and blast wave phenomena.

Table 3. Distribution of wounded casualties by primary threat delivery platform / weapon system [11]

Threat category		Percentage
Gunshot fire		4
Fragment and blast injuries from	artillery shells	20.5
	UAV	75.5

This distinction is further reflected in Table 4, which categorizes casualties strictly by terminal biophysical trauma mechanism. Unlike Table 3, the data for Tables 4, 5 and 6 originates from forward medical battalions (Level 2 care) evaluating 2014 casualties during the active operational phase and 1186 during the inactive phase. The prevalence of these injury mechanisms across different phases of hostilities is presented in Table 4, while their corresponding anatomical distribution is shown in Table 5. Across both phases, fragment impacts (66.5%–79.3%) and blast trauma (5.2%–15.1%) represented the primary biophysical mechanisms of wound creation.

Table 4. Prevalence of combat casualties categorized by terminal injury mechanism across operational phases. [11]

Mechanism of injury	Active phase	Inactive phase
Bullet wounds	5.6	2.2
Fragment injuries	79.3	66.5
Blast injuries	15.1	5.2
Other	/	26.1

Penetrating cavity and deep injuries, including neck wounds, were observed in 69% of cases during the active phase and 23.8% during the inactive phase of hostilities (Table 6). The proportion of lightly wounded patients increased markedly from 22.4% to 72.3%, reflecting the positional, low-intensity nature of current hostilities and the reduction of heavy enemy firepower. This trend also corresponds with the notable decrease in penetrating injuries, from 69% during the active phase to 23.8% during the inactive phase.

Table 5. Anatomical distribution of dominant injuries across operational phases [11]

Localisation	Active phase	Inactive phase
Head	4.5	11
Neck	2	0.7
Chest	6.9	8.9
Abdomen	5.2	5.9
Extremity	81.4	58.2

Table 6. Incidence of penetrating injuries [11]

Localisation	Active phase	Inactive phase
Head	58.2	25.9
Neck	56.1	25.9
Chest	75.2	18.1
Abdomen	75.2	27.7

Extremity injuries remain the most common among Russian casualties, with frequencies of 81.4% and 58.2% across different phases. In the inactive phase, 917 casualties (77.3%) presented with extremity injuries, though severe cases involving extensive soft tissue loss, amputations, or long bone fractures have decreased to 10.1%, compared to up to 60% during the active phase. Combined injuries without a dominant lesion rank second (14.5%), with 90.1% classified as mild. Head injuries constitute 11% of cases, of which only 21.4% involve TBI, primarily affecting the unprotected facial region, including the eyes. Thoracic injuries also deserve specific attention. The proportion of penetrating chest injuries has decreased to 18.1%, compared to 75.2% at the beginning of hostilities.

3. PERFORMANCE OF BODY ARMOUR

A comparative analysis of combat-related injuries in the conflicts in Iraq and Afghanistan (2001–2017) and the ongoing Russo-Ukrainian war (2022–2026) provides a functional framework for evaluating the operational performance and limitations of modern personal ballistic protection equipment. The observed injury patterns, together with their temporal and spatial shifts, directly reflect both the protective effectiveness of body armour and the evolving nature of battlefield threats.

Across both conflict periods, extremity injuries remain the most prevalent category, accounting for approximately 50–52% of casualties in Iraq and Afghanistan and reaching up to 81.4% among Russian forces, with similar trends reported among Ukrainian personnel. This persistent dominance highlights the inherent limitation of body armour systems, which primarily protect the central torso while leaving the limbs exposed (Figure 2).

Head, face, and neck injuries show an increasing relative prevalence over time. In Iraq and Afghanistan, head and neck trauma constituted approximately 28% of injuries, with an increasing trend during later stages of conflict. In the Russo-Ukrainian war, these injury types, including TBI, remain prominent and are further amplified by the increased use of UAV-delivered payloads and thermobaric weapons. This shift highlights the impact of modern weaponry and the limitations of protective equipment in shielding these regions.

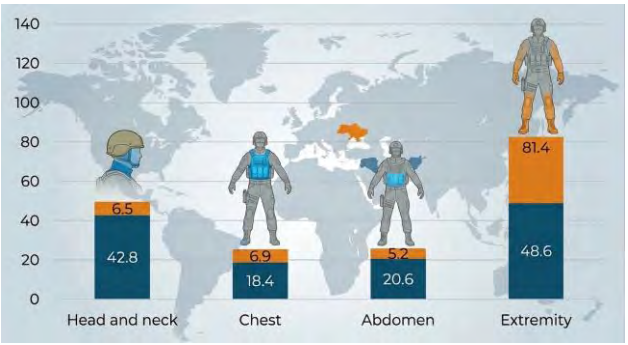


Figure 2. Change in anatomical distribution of injuries

Explosive mechanisms represent the dominant cause of injury in both conflict environments, with an increased impact in the recent war (Figure 3). In Iraq and Afghanistan, explosives accounted for approximately 74% of combat casualties, while in the Russo-Ukrainian conflict, they contribute to a majority of injuries across both sides. Fragments and blast induced injuries not only produce a higher proportion of multi-region and penetrating injuries but also result in more severe trauma, including burns, pulmonary injury, and TBI.

The distribution of thoracic and abdominal injuries demonstrates the effectiveness of modern ballistic protection. In Iraq and Afghanistan, these injuries accounted for approximately 10% of casualties, while in Ukraine, the proportion of penetrating thoracic injuries has decreased markedly from 75.2% during the active phase to 18.1% in later phases. This confirms that modern body armour systems, combining soft and hard ballistic elements, are highly effective in mitigating lethal injuries caused by rifle rounds and fragmentation. Ceramic plate systems significantly reduce mortality by preventing penetration of the thorax and upper abdomen, with remaining injuries often associated with coverage gaps or edge impacts. Historical data from explosive incidents, including IEDs and mines, further confirm that torso injuries are often mitigated in severity due to ballistic protection [12].

Figure 3. Change in the mechanism of injuries

Despite this effectiveness, current body armour systems exhibit critical limitations. Most importantly, they do not protect against blast overpressure, which has become a defining injury mechanism in modern warfare. Traumatic brain injury, particularly blast-induced TBI, is now considered a signature injury of modern warfare. While armour effectively protects against fragments, it does not mitigate the transmission of pressure waves, which can cause significant neurological damage even in the absence of external head penetration.

The protection offered by current systems is achieved at the cost of increased weight and reduced mobility. Body armour is heavy, thermally burdensome, and ergonomically restrictive, leading to decreased soldier mobility, increased fatigue, and impaired physical and cognitive performance. In hot operational environments, heat retention further increases the risk of thermal stress, thereby limiting sustained operational effectiveness.

Despite advances in protective design, significant portions of the body still remain unprotected. Regions such as the head, face, neck, groin, lateral torso, and extremities are particularly vulnerable, especially under high-energy ballistic and explosive conditions. While soft armour protects against lower-energy fragments and pistol rounds, it is insufficient against rifle projectiles. Hard armour systems, although effective, are limited to coverage of the central torso due to weight constraints, making full-body protection impractical without the use of emerging technologies such as exoskeletons, which remain in developmental stages with potential long-term applicability.

The protective effect of body armour has also influenced injury epidemiology. While body armor significantly reduces the incidence of fatal torso trauma, it does not lead to a proportional reduction in overall casualty rates. Instead, the introduction of advanced body armor alters the observed epidemiology of operational injuries: casualties who previously would have succumbed to lethal torso injuries now survive, thereby shifting the distribution of recorded non-fatal injuries toward unprotected anatomical regions (such as the head, neck, and extremities). Consequently, the apparent increase in non-torso injuries reported in combat databases [7] reflects a survivability bias, driven by improved survival from otherwise fatal torso trauma and exposure to modern threats such as explosive devices, urban combat, and targeted fires, rather than an inherent vulnerability induced by the armor itself.

4. NOVEL BALLISTIC PROTECTION MATERIALS

The development of materials for ballistic protection has evolved significantly in recent decades, driven by the need to simultaneously address multiple threats, including high-velocity projectile penetration, blast loading, and stab or spike resistance. Modern protective systems are no longer based on a single material solution but rely on hybrid, multifunctional architectures that combine different classes of materials to optimize energy absorption, dissipation, and structural integrity. Recent research trends emphasize lightweight, flexible, and adaptive materials, particularly those enabled by nanotechnology, advanced polymers, and shear thickening fluids and gels.

4.1 High-performance fibers and polymer composites

High-performance fibers remain the foundation of modern soft and hybrid armor systems. Materials such as aramid fibers and ultra-high molecular weight polyethylene (UHMWPE) exhibit exceptional specific strength and energy absorption capacity, making them indispensable for ballistic protection [13].

Recent advances have focused on improving interfacial properties and structural organization within composite systems. For instance, SiC nanoparticle-modified UHMWPE/polyurethane composites have demonstrated enhanced ballistic limits due to improved load transfer and reduced interlayer slippage [14]. Similarly, multilayer polymer fiber composites have been optimized to maximize energy absorption through controlled delamination and progressive failure [15].

Polymer matrix modifications, including epoxy toughening and thermoplastic reinforcement, contribute to improved impact resistance and structural stability [16].

In addition to projectile resistance, these materials also contribute to blast mitigation by distributing stress waves over larger areas and reducing backface deformation.

4.2 Shear thickening fluids and gels

One of the most promising developments in recent years is the integration of shear thickening fluids (STFs) and shear thickening gels (STGs) into textile and composite systems. STFs and STGs are non-Newtonian materials that exhibit a dramatic increase in viscosity under high shear rates, enabling dynamic stiffening during impact events. This property allows flexible materials to behave as rigid protective barriers upon loading.

Extensive research demonstrates that STF-impregnated fabrics significantly enhance both ballistic and stab resistance. The increased inter-yarn friction and restricted fiber mobility lead to improved energy absorption and reduced penetration depth [17, 18]. Furthermore, the incorporation of nanoparticles such as silica, graphene oxide, or clay into STFs further enhances their rheological performance and protective capabilities [19].

Recent innovations include hybrid systems combining STFs with ceramic particles such as boron carbide (B_4C), which provide additional hardness and penetration resistance [20]. These systems are particularly effective against sharp objects and low-velocity threats, where traditional fiber systems may fail.

Beyond ballistic protection, STFs also show significant potential in blast mitigation. Their ability to absorb and dissipate energy through reversible phase transitions contributes to reduced transmitted forces and improved wearer safety [21].

4.3 Nanocomposites

Nanotechnology has emerged as a transformative approach in the design of advanced ballistic materials. Polymer nanocomposites, reinforced with nanoparticles such as carbon nanotubes (CNTs), graphene, MXene, or ceramic nanofillers, exhibit significantly improved mechanical properties, including stiffness, toughness, and energy absorption capacity [22, 23].

The primary advantage of nanocomposites lies in their ability to modify material behavior at the molecular and interfacial level. For example, CNT-reinforced UHMWPE fabrics demonstrate enhanced impact resistance due to improved load distribution and crack bridging mechanisms [24]. Similarly, graphene-based nanocomposites offer exceptional strength-to-weight ratios and improved thermal and electrical properties, which can be beneficial for multifunctional protective systems.

The addition of hard nanoparticles increases surface hardness and reduces local deformation, thereby enhancing stab resistance [25]. Furthermore, nanostructured coatings and interlayers can significantly improve interfacial bonding and prevent delamination under dynamic loading.

Despite their advantages, challenges remain in terms of scalability, cost, and uniform dispersion of nanoparticles. Nevertheless, ongoing research continues to address these limitations, making nanocomposites a key area in ballistic protection.

4.4 Ceramic materials and hybrid armor systems

Ceramic materials such as alumina, silicon carbide (SiC), and boron carbide (B_4C) are widely used in hard armor systems due to their high hardness and compressive strength. Their primary function is to disrupt and erode incoming projectiles, thereby reducing their penetration capability. Recent studies highlight the superior performance of boron carbide in terms of hardness and low density, making it particularly attractive for lightweight armor applications [26].

Ceramics are inherently brittle, which limits their ability to absorb energy and makes them susceptible to cracking under repeated impacts. To overcome these limitations, modern armor systems employ hybrid configurations, combining ceramic front layers with fiber-reinforced composite backings. This approach allows efficient projectile fragmentation by the ceramic layer, followed by energy absorption and distribution by the composite layers [27].

Functionally graded materials (FGMs) represent an emerging concept in this domain. By gradually varying material properties across the thickness, FGMs can reduce stress concentrations and improve overall impact resistance [28]. Such materials are particularly promising for multi-threat protection, as they can be tailored to respond differently to various loading conditions.

4.5 Lightweight and multifunctional materials

The demand for lightweight and multifunctional protective systems has led to the exploration of novel materials such as aerogels, sandwich composites, and bio-inspired structures. Silica aerogels, for example, offer

extremely low density and high energy absorption capacity, making them suitable for impact mitigation and thermal protection [31].

Sandwich structures, combining different materials in layered configurations, enable optimization of mechanical performance by exploiting the synergistic effects of each component. These systems can be tailored to address specific threats, including ballistic impact, blast loading, and stab resistance.

Bio-inspired materials represent another promising direction [23]. Inspired by natural protective systems such as nacre or armor shells, these materials utilize hierarchical structures to achieve exceptional toughness and energy dissipation. Such designs have shown significant potential in improving impact resistance while maintaining low weight.

5. TEST STANDARDS LIMITATIONS

Standards for testing personal protective ballistic equipment, such as NIJ standards in the United States and GOST standards in Russia, have been developed for different operational conditions, which leads to differences in the definition of protection levels, threat types, and testing methods. One of the key limitations of the NIJ 0101 family of standards relates to Level IV protection, specifically the issue that certain ammunition (e.g., 7.62 mm with an M993 projectile) has greater penetration capability than the reference round used in the standard. It is important to emphasize that the NIJ standard was primarily developed for law enforcement use, where threats are statistically limited to handgun and standard rifle calibers, whereas military battlefields are characterized by the presence of a wide variety of ammunition calibers and projectile types.

In contrast, in the Russian Federation, the GOST 34286-2017 standard recognizes the need for higher protection levels corresponding to military conditions. The introduction of Level BR6 represents a response to the emergence of extremely high-penetration ammunition, where testing is conducted using 12.7×108 mm caliber ammunition from a distance of 50 meters. However, this standard only considers the presence or absence of penetration, without addressing the degree of trauma to the human body.

The most pronounced problems are observed in helmet testing. The NIJ 0106.01 standard has not been revised since 1980, making it insufficiently adapted to modern design solutions and threats. Over the past decades, there has been significant development in helmet design, particularly high-cut variants that provide better compatibility with communication and optical equipment but simultaneously reduce the protective surface area. Such helmets also feature a greater number of shell openings for mounting equipment. The standard does not account for these changes in geometry, which leads to problems in uniform testing. Due to the outdated nature of the NIJ 0106.01 standard, in practice, ammunition for Level IIIA from the NIJ 0101 standard family is often used for testing.

An additional limitation in ballistic helmet evaluation is the small number of standards designed to quantify non-penetrating neurotrauma. Modern combat helmets reliably arrest small-arms projectiles and fragments; however, energy transfer and shock wave propagation remain critical vectors for Traumatic Brain Injury (TBI). In combat environments, TBI represents a highly heterogeneous spectrum driven by distinct physical mechanisms:

- Penetrating TBI, resulting from direct projectile or fragment perforation;
- Blunt impact TBI, caused by Behind-Helmet Blunt Trauma (BHBT), where the severe dynamic backface deformation of a non-perforated helmet shell strikes the skull directly;
- Acceleration-induced TBI, driven by rapid linear and rotational accelerations imparted to the head, inducing shear stresses and diffuse axonal injury within brain tissue even in the absence of direct cranial impact; and
- Blast-associated TBI, induced by primary blast overpressure wave transmission through the helmet structure without mechanical shell compromise.

Existing ballistic standards (such as VPAM HVN 2009 and HPW-TP-0401.01B) attempt to evaluate head trauma via blunt impact metrics, but they are neither universally mandated nor capable of assessing rotational kinematics or blast wave dynamics. The absence of a standardized framework for quantifying non-penetrating TBI risk, specifically acceleration and blast overpressure mitigation, means helmets can fully satisfy conventional ballistic penetration criteria while failing to protect personnel from non-penetrating neurological trauma.

The greatest problem arises in blast-related injuries, as there is no standard for assessing protection. Blast waves represent a dominant injury mechanism in modern conflicts, yet their effects—including wave transmission through materials and indirect impacts on the central nervous system—are not considered in certification procedures for personal protective ballistic equipment.

All of these facts indicate a fundamental gap between laboratory standards and real combat conditions. While existing standards are successful in defining minimum ballistic protection criteria, they fail to encompass the complexity of modern injuries, particularly those caused by explosions and indirect biomechanical effects. This mismatch represents one of the key challenges in the further development of protective ballistic equipment and testing standards.

6. CONCLUSION

The analysis conducted in this paper unequivocally confirms that the development of personal ballistic protective equipment must keep pace with the dynamic changes on the modern battlefield. It has been established that dominant threats, such as drones and precision artillery, create complex injury patterns that exceed the protective capacities defined by outdated standards, particularly regarding head and facial protection.

The key contribution of this work lies in identifying the potential of advanced materials, such as carbon nanotube-reinforced composites and shear-thickening fluids, which enable the development of lightweight and flexible equipment resistant to multiple threats. It is concluded that an urgent revision of testing standards is essential for the future effectiveness of personal ballistic protective equipment, alongside the integration of modular and smart systems that balance maximum ballistic resistance with mobility.

References

[1] MILORADOVIĆ,N.: Contemporary security challenges and their impact on the increase of the military spending and the budget for the development and procurement of weapons and military equipment, Vojno delo, 75(3) (2023) 30–47.

[2] KUNERTOVA,D.: *Drones have boots: Learning from Russia's war in Ukraine*, Contemporary Security Policy, 44(4) (2023) 576–91.

[3] WATLING,J., REYNOLDS,N.: Tactical Developments During the Third Year of the Russo–Ukrainian War, RUSI, London, 2025.

[4] BELMONT,P., McRISKIN,B.J., SIEG,R.N., BURKS,R., SCHOENFELD,A.,: *Combat wounds in Iraq and Afghanistan from 2005-2009*, Journal of Trauma and Acute Care Surgery, 73(1) (2012) 3-12.

[5] ESKRIDGE,S.L., MACERA,C.A., GALARNEAU,M.R., HOLBROOK,T.L., WOODRUFF,S.I., McGREGOR,A.J., MORTON,D.J., SHAFFER,R.A.: *Injuries from combat explosions in Iraq: Injury type, location, and severity*, Injury, 43(10) (2012) 1678–82.

[6] HOENCAMP,R., VERMETTEN,E., TAN,E.C.T.H., PUTTER,H., LEENEN,L.P.H., HAMMING,J.F.: Systematic review of the prevalence and characteristics of battle casualties from NATO coalition forces in Iraq and Afghanistan, Injury, 45(7) (2014) 1028–34.

[7] TONG,D., BEIRNE,R.: Combat Body Armor and Injuries to the Head, Face, and Neck Region: A Systematic Review, Military Medicine, 178(4) (2013) 421–426.

[8] HOWARD,J.T., KOTWAL,R.S., TURNER,C.A., JANAK,J.C., MAZUCHOWSKI,E.L., BUTLER,F.K., et al., Use of Combat Casualty Care Data to Assess the US Military Trauma System During the Afghanistan and Iraq Conflicts, 2001-2017, JAMA Surgery, 154(7) (2019) 600-8.

[9] KHORRAM-MANESH,A., GONIEWICZ,K., BURKLE,F.M., ROBINSON,Y.: *Review of Military Casualties in Modern Conflicts—The Re-emergence of Casualties From Armored Warfare*, Military Medicine, 187(3–4) (2022) 313–21.

[10] LAWRY,L.L., MANI,V., HAMM,T.E., JANVRIN,M., JUMAN,L., et al: Qualitative assessment of combat-related injury patterns and injury prevention in Ukraine since the Russian invasion, BMJ Military Health, 172(2) (2026) 154–9.

[11] KASIMOV,R.R., HEZHEV,M.A., SAMOKHVALOV,I.M., ZAVRAZHNOV,A.A., KOVALENKO,S.A., PROSVETOV,V.A.: *Impact of combat conditions on casualties and surgical care*, Bulletin of the Russian Military Medical Academy, 27(4) (2025) 517–26.

[12] PELEG,K., RIVKIND,A., AHARONSON-DANIEL,L.: *Does Body Armor Protect from Firearm Injuries?,* Journal of the American College of Surgeons, 202(4) (2006) 643–8.

[13] JOSHI,A., MISHRA,A., SAXENA,V.K.: Impact response and energy absorption mechanisms of UHMWPE fabric and composites in ballistic applications: A comprehensive review, Composites Part A: Applied Science and Manufacturing, 185 (2024).

[14] QI,C., ZHAO,Y., WANG,Z.: Interfacial decoration of UHMWPE composites with SiC nanoparticles to improve their ballistic limitations, Polymer Composites, 46(S1) 2025.

[15] PIZZATTO,F., PIZZATTO,S.M.S., OLIVEIRA,A.P.N., LIMA,E.S., NASCIMENTO,L.F.C., MONTEDO,O.R.K., ARCARO,S.: *Ballistic Performance of Multilayer Polymer Fiber Composite Plates*, Polymers for Advanced Technologies, 36(9) (2025).

[16] B. HOU, T. LAN, L. ZU, S. DONG, AND L. XU: *Research status of mechanical modification of epoxy resin*, Polymer Bulletin, 82(16) (2025) 11323–56.

[17] LIU,S., ZHAO,C., WEN,Z.: *Impact protection performance of Kevlar fabric composites impregnated with modified multiphase shear thickening fluids*, Thin-Walled Structures, 214 (2025).

[18] WANG,J., FENG,B., WANG,Z., CAO,M., CHEN,L.: *Review: Research progress on the impact protection effect of shear thickening materials under dynamic loading*, Journal of Materials Science, 60(38) (2025) 17371–414.

[19] ASTARAKI,S., EASAVINEJAD,H., HASANNEZHAD,H., PAYAMI,S.R., BAGHERI,Y., NOURBAKHSH,S.H., KHABAZ-AGHDAM,A.: *Enhanced rheological performance of shear thickening fluids: effects of graphene oxide nanoplatelets and clay nanoparticles*, Colloid and Polymer Science, 303(5) (2025) 907–21.

[20] AŞKAN,A., AYDIN,M.: *Enhancing the Impact and Stab Resistance of Aramid Fabrics Via Boron Carbide-Reinforced Shear Thickening Suspensions*, Applied Composite Materials, 32(6) (2025) 2475–96.

[21] SILVA,D.S., MARQUES,M.D.F.V., LEMOS,M.F., MONTEIRO,S.N.: *Shear thickening fluids and gels as reinforcement in ballistic vest: an update review*, Journal of Materials Research and Technology, 40 (2026) 547–76.

[22] SELIM,M.S., EL-SAFTY,S.A., SHENASHEN,M.A., ELMARAKBI,A., *Advances in polymer/inorganic nanocomposite fabrics for lightweight and high-strength armor and ballistic-proof materials*, Chemical Engineering Journal, 493 (2024).

[23] ABTEW,M.A., DEJENE,B.K., ATALIE,D. BAJUA,M., *Advanced surface functionalization for smart, flexible and multi-threat impact resistance protective textiles and composites*, Defence Technology, (2026).

[24] LU,C., JI,W., LU,X., XU,C., ZHANG,J., WANG,X. BA,S.: *Improving the anti-impact performance of ultra-high molecular weight polyethylene fabric intercalated with carbon nanotubes and shear thickening fluid*, Polymer Composites, 46(6) (2025) 5255–66.

[25] SIDDIQUI,M.A.S., RABBI,M.S., RAHMAN,M.D.T.: *An experimental review on enhancing the low and high-velocity impact resistance of NFCs through nanoparticle reinforcement*, Journal of Reinforced Plastics and Composites, 44(17-18) (2025)1270–98.

[26] MAKAOUI,R., MEHELLI,O., TRIA,D.E., HEMMOUCHE,L., HABES,A., DERRADJI,D.M.: *The impact of boron carbide in soft and hard body armor: A comprehensive evaluation on Kevlar and UHMWPE fabrics*, Polymer Composites, 45(12) (2024) 11062–75.

[27] MUTU,H. B., ÖZER,A.: *Experimental and finite element analysis of ballistic properties of composite armor made of alumina, carbon and UHMWPE*, Polymer Composites, 45(15) (2024) 13844–860.

[28] PATEL,M.: *A comprehensive review of functionally graded materials and their ballistic impact performance: Current status and future challenges*, Next Materials, 8 (2025).

[29] WANG,L., KONG,J., CHU,D., WANG,Y., WANG,T., LIU,Z.: *Enhancing ballistic performance: Effect of polyurea coating on backface deformation of UHMWPE laminates*, Composite Structures, 355 (2025).

[30] WANG,Z., YANG,L., YANG,S., LIU,P., WU,L.: *A novel composite material of aramid fiber-reinforced polyurea elastomer matrix: mechanical properties and ballistic performance*, Composites Science and Technology, 271 (2025).

[31] KUMAR,S. L.P., DEEBAN,B., KUMAR,D.S., BARANITHARAN,P., SRIARIYANUN,M., GOPAL,M.: *Unveiling the armor: experimental study on ballistic impact resistance of silica aerogel, sorbothane, and graphite sandwich composites*, Discover Materials, 5(1) (2025) 224-49.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CARBON-BASED ADSORPTIVE MATERIALS FOR MITIGATION OF ORGANOPHOSPHATE CONTAMINANTS IN AQUATIC SYSTEMS

VLADAN J. ANIĆIJEVIĆ
Military Academy, University of Defense in Belgrade, Serbia, anicijevicj.v@gmail.com

KATARINA KOKANOV STANKOVIĆ
Faculty of Physical Chemistry, University of Belgrade, Serbia, 20240319@student.ffh.bg.ac.rs

VEDRAN MILANKOVIĆ
Department of Physical Chemistry, "VINČA" Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, vedran.milankovic@vin.bg.ac.rs

TAMARA TERZIĆ
Department of Physical Chemistry, "VINČA" Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, tamara.tasic@vin.bg.ac.rs

RADOVAN M. KARKALIĆ
Military Academy, University of Defense in Belgrade, Belgrade, rkarkalic@yahoo.com

TAMARA LAZAREVIĆ-PAŠTI
Department of Physical Chemistry, "VINČA" Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, tamara@vin.bg.ac.rs

Abstract: *Rapid industrialization and technological advancement have significantly increased the presence of hazardous chemicals in environmental matrices, posing risks not only to public health and ecosystems but also to environmental and water security. Among these contaminants, organophosphate compounds represent a critical class of toxic substances. While primarily used as agricultural pesticides, organophosphates share structural and toxicological features with certain nerve agents, making their environmental persistence and potential misuse a matter of concern within CBRN (Chemical, Biological, Radiological, and Nuclear) risk frameworks. Aquatic systems are particularly vulnerable, as wastewater discharges facilitate the transport of pesticides, synthetic dyes, and heavy metals into rivers, lakes, and groundwater reservoirs. The development of efficient, scalable, and low-cost remediation materials is therefore essential for both environmental protection and strategic water safety management. This study investigates the adsorption of chlorpyrifos onto bio-derived carbon materials synthesized via controlled carbonization of waste biomass. The objective was to assess their removal efficiency and evaluate their potential applicability in defensive water purification systems designed for contamination response and preventive risk mitigation. Material characterization was conducted using scanning electron microscopy (SEM), Fourier-transform infrared spectroscopy (FTIR), and Brunauer-Emmett-Teller (BET) surface area analysis to determine morphology, surface chemistry, and porosity. Adsorption experiments were monitored by ultra-performance liquid chromatography (UPLC) at three operational temperatures (22 °C, 30 °C, and 35 °C). Adsorption efficiency and equilibrium time were determined to assess operational feasibility under variable environmental conditions. The results provide mechanistic insight into interactions between organophosphate compounds and carbon-based materials and support the development of bio-derived adsorbents as cost-effective components in resilient water treatment systems relevant to environmental security and CBRN preparedness.*

Keywords: *chlorpyrifos, nerve agents, bio-derived carbon materials, adsorption.*

1. INTRODUCTION

Surface water pollution is a growing environmental problem that significantly affects ecosystems, human health, and biodiversity. Organophosphate pesticides, which are primarily used in agriculture to control pests and increase yields, stand out among the many pollutants degrading water quality. They do not decompose spontaneously in nature, adversely affect aquatic ecosystems, and are highly toxic to living organisms [1]. They reach the waters through runoff from agricultural lands, where they can remain for a long time, and through interactions with other substances in the water, more toxic compounds can form. Exposure to organophosphate pesticides is associated with many health diseases, both in humans and animals [2].

Corresponding author: V. Anićijević;
E-mail: anicijevicj.v@gmail.com

DOI: 10.5937/OTEH26109A

The aim of this work was to investigate the adsorption of chlorpyrifos (CHP) on carbon materials obtained from biowaste. The adsorption of these substances was investigated at different concentrations, three temperatures, and various contact times, with a constant amount of material. A carbon material derived from biowaste, obtained by carbonizing used rabbit litter, was used as an adsorbent. Morphological characterization of the material was performed using a scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDX). Functional surface groups were identified using Fourier transform infrared spectroscopy (FTIR), and the material's specific surface area was also examined using BET (Brunauer-Emmett-Teller) analysis.

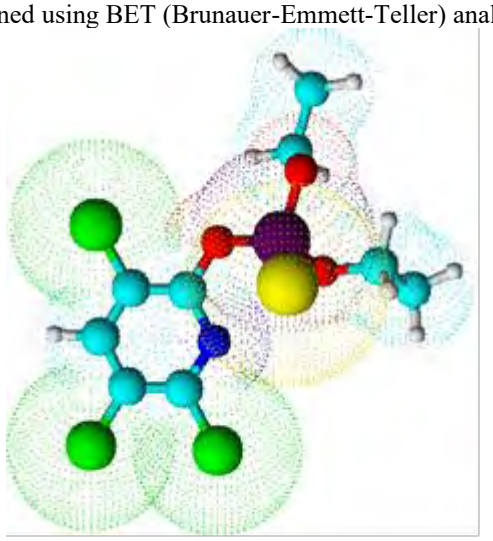


Figure 1. Chemical structure of CHP

CHP (Figure 2) adsorption on carbon materials was monitored experimentally using ultra-high-performance liquid chromatography (UPLC). Contaminant and carbon material suspensions were incubated for a specified period of time at three different temperatures, after which the time required for equilibrium to be established will be determined, after which no further CHP adsorption on the adsorbent will occur. Adsorption efficiency will be tested at a constant carbon material concentration.

The results will be processed using the Dubinin-Radushkevich and Dubinin-Astakhov isotherm models to determine characteristic adsorption parameters and describe the adsorbent surfaces.

2. EXPERIMENTAL

2.1. Chemicals

Daily prepared working solutions of CHP in ethanol were used in the work, at concentrations ranging from 5×10^{-4} to 5×10^{-6} mol dm^{-3}, obtained by diluting the 1×10^{-1} mol dm^{-3} solution. Ethanol (C_2H_5OH) and acetonitrile (CH_3CN), HPLC grade, were obtained from J.T.Baker.

2.2. Methods

UPLC – Chromatographic analyses were performed on an ultra-high-performance liquid chromatography Waters ACQUITY system with a binary pump (Binary Solvent Manager), an automatic sampler (Sample Manager), and a detector (Photo Diode Array - PDA) in the wavelength range from 190 to 500 nm. Analyses were performed at a constant eluent composition (isocratic conditions). An ACQUITY UPLC TM BEH C18 chromatographic column was used as the stationary phase, and the mobile phase was 20% water, 80% acetonitrile.

Other laboratory equipment – included a Sartorius 2632 analytical balance and a Mettler technical balance. Eppendorf Research automatic pipettes were used to measure liquid samples. Samples were mixed on a Vortex V1 plus BOECO mixer and an Orbital shaker-Incubator Grant-bio ES-20 mixer at a constant, set temperature. Separation of the solid and liquid phases was performed using a Mini Spin PLUS centrifuge (Eppendorf) at 14,500 rpm. The samples were filtered using a KX Syringe Filter with a 13 mm diameter and a 0.22 µm pore size.

3. RESULTS AND DISCUSSION

3.1. Synthesis of carbon materials derived from biowaste

The carbon materials (CM) used as adsorbents were obtained from biowaste, by carbonizing the used rabbit litter. The first adsorbent includes all components of the spent litter (CM I), while the second contains unused sawdust (CM II). The first step in the synthesis process is to dry the biowaste in an oven at 90 °C for 24 h, after which they were carbonized in a furnace atmosphere at 650 °C for 1 h. At the end, carbon material suspensions were prepared in ethanol at a concentration of 2 mg cm^{-3} (suspension stock).

3.2. Characterization of carbon materials

SEM and EDX spectroscopy – The morphology of the carbon materials was analyzed using a SEM (JEOL, Ltd., Akishima, Tokyo, Japan). The homogenized sample powder was placed on a carbon support strip using tweezers, then introduced into the SEM chamber. The chemical composition of the investigated carbon materials was determined using EDX.

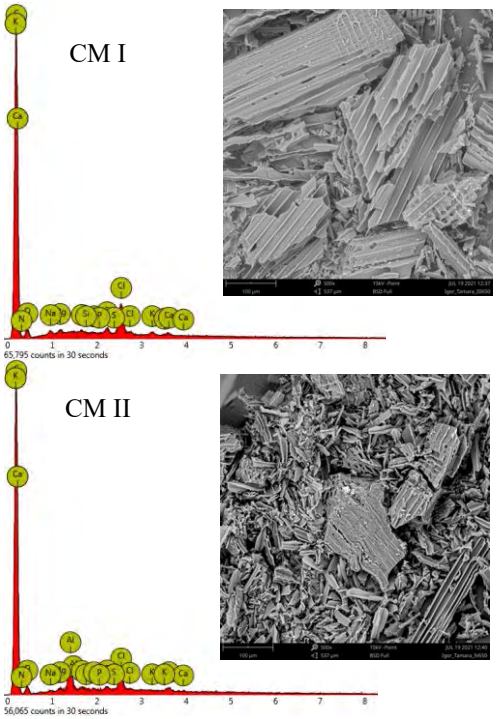


Figure 3. SEM micrograph and EDX spectrum of CM

From the micrographs shown in Figure 2, it can be seen that the primary structure of both materials is retained, and that they are rough and porous. Based on the spectra, the most abundant element is carbon, at 92.94% in CM I and 93.02% in CM II. In addition, the composition of CM I includes 5.26% oxygen and 1% chlorine. CM II also consists of oxygen with 4.59% and aluminum with 1.16%. Other elements, such as calcium, magnesium, silicon, and others, are present in trace amounts, that is, at levels below 1%.

BET analysis – In order to determine the specific surface area of the carbon material, an automatic volumetric sorption analyzer (Sorptomatic 1990 Finnigan) was used. The analysis was performed using isothermal adsorption of nitrogen at 77.13 K. Before adsorption, the samples were degassed in a vacuum for 4 h at 90 °C and 16 h at 200 °C.

The value of the specific surface for CM I is S_{BET} = 10 m^2g^{-1} while for CM II, S_{BET} = 487 m^2g^{-1}. It is observed that the second material has almost 50 times greater specific surface area compared to the first.

FTIR – In order to determine the characteristic functional groups present in the adsorbents, a Nicolet iS20 FT-IR spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) was used. The wavelength range was from 4000 to 500 cm^{-1}, with 64 scan/a every 4 cm^{-1}.

The band assignments were made in accordance with the literature [3].

Each functional group has a characteristic position on the FTIR spectrum (Table 1). Each band present corresponds to a specific vibration of the particular chemical bond of the given material.

Table 2. Characteristic functional groups of CM I and CM II

Band position (cm^{-1})	Functional group	Type of vibration
617	O-H free	deformation
655	-COOH	deformation
702	-COOH	deformation
863	C-H aromatic out-of-plane	bending
887	C-H aromatic out-of-plane	bending
1033	C-H aromatic compounds	deformation
1110	C-O	deformation
1380	C-N	stretching
1458	C-H skeletal vibration aromatic ring	deformation
1589	C=C skeletal vibration of aromatic ring	stretching

CM I and CM II have common peaks originating from the C-H bending vibration, the -COOH deformation vibration and the O-H free, deformation vibration. However, in the FTIR spectrum of CM II, a bathochromic shift of the C-H and -COOH bands can be observed. The cause of this phenomenon may lie in the different chemical compositions of these two materials. Despite the shift of the two mentioned characteristic bands, the O-H vibration occurs in both materials at the same position. In addition, CM I contains a band characteristic of the C=C stretching vibration, which is not present in the spectrum of CM II. CM II is also characterized by various deformation vibrations, including C-H and C-O, and by one C-N stretching vibration.

3.3. Adsorption of pesticides on carbon materials at different temperatures

The adsorption of CHP was investigated at 22 °C, 30 °C, and 35 °C, with a constant amount of adsorbents, over a pesticide concentration range of 5×10^{-4} to 1×10^{-6} mol dm^{-3}. The experimental data were fitted using the nonlinear forms of the Dubinin-Radushkevich and Dubinin-Astakhov isotherm models, as shown in Figure 4. The corresponding adsorption parameters, which provide insight into the surface characteristics and adsorption behavior of the adsorbents, are summarized in Table 3.

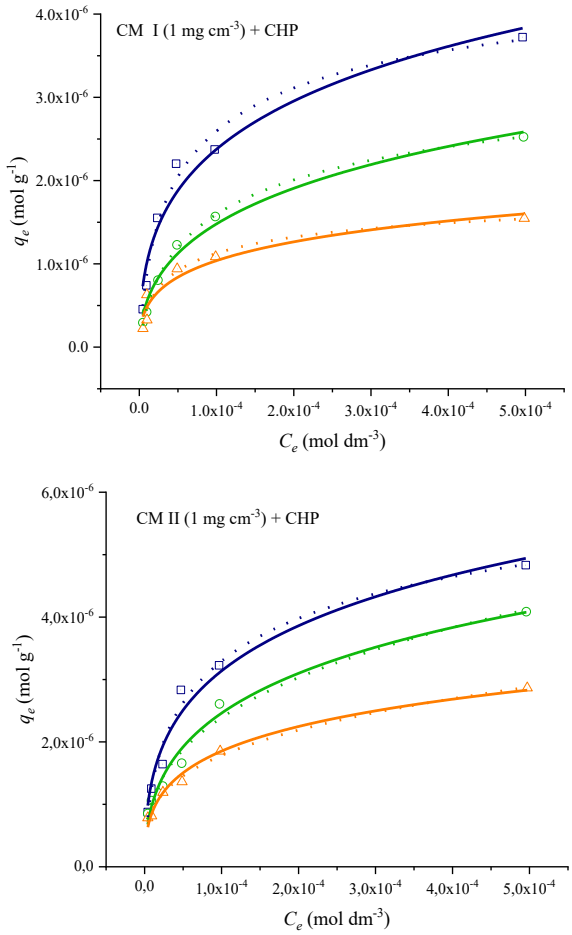


Figure 5. Nonlinear forms of isotherms applied to the adsorption process of CHP on CM:

□ 22 °C ○ 30 °C △ 35 °C

Dubinin-Radushkevich · · · Dubinin-Astakhov

The obtained results for the adsorption of CHP on both carbon materials indicate that the process can be well described by the Dubinin-Radushkevich and Dubinin-Astakhov isotherm models at all temperatures (Figure 6). It can be observed that CM II exhibits a higher adsorption capacity for CHP compared to CM I.

Table 4. Parameters of adsorption isotherm models for adsorption, obtained by fitting experimental data

T/°C	*22*	*30*	*35*
Dubinin – Radushkevich *CM I*			
q_{DR}/mol g^{-1} ×10^6	11 ± 2	8,5 ± 0,9	4,0 ± 0,8
K_{DR}/mol^2 kJ^{-2}	260 ± 30	300 ± 20	230 ± 40
E_a/J mol^{-1}	44 ± 3	41 ± 2	46 ± 4
χ^2×10^{14}	7,23928	1,21053	2,10462
R^2	0,95958	0,98599	0,9324
Dubinin – Radushkevich *CM II*			
q_{DR}/mol g^{-1} ×10^6	13 ± 2	12 ± 2	7,0 ± 0,6
K_{DR}/mol^2 kJ^{-2}	240 ± 20	270 ± 20	230 ± 10
E_a/J mol^{-1}	45 ± 2	43 ± 2	46 ± 2
χ^2×10^{14}	6,55679	3,69693	1,05219
R^2	0,97612	0,98031	0,98633
Dubinin – Astakhov *CM I*			
q_{DA}/mol g^{-1} ×10^6	4,6 ± 0,7	3,8 ± 0,3	1,9 ± 0,6
n	5 ± 2	4 ± 0,4	5 ± 2
E_a/J mol^{-1}	26 ± 1	23,8 ± 0,6	26 ± 2
χ^2×10^{14}	3,27003	16,6068	1,87611
R^2	0,98631	0,99856	0,9548
Dubinin – Astakhov *CM II*			
q_{DA}/mol g^{-1} ×10^6	7 ± 2	32 ± 9	40 ± 10
n	3 ± 1	1 ± 1	0,9 ± 0,8
E_a/J mol^{-1}	24 ± 2	10 ± 2	6 ± 5
χ^2×10^{14}	5,68135	4,29678	8,0718
R^2	0,98448	0,98283	0,99213

Based on the q_{max} values derived from the isotherm parameters, the maximum amount of adsorbed CHP decreases with increasing temperature (Table 5). The Dubinin-Radushkevich model indicates that the adsorption is predominantly physical (E < 8 kJ mol^{-1}), while the Dubinin-Astakhov model provides a better fit to the experimental data, likely due to surface heterogeneity of the adsorbents (n ≠ 2) [4].

4. CONCLUSION

In this study, the adsorption of CHP onto carbon materials derived from biowaste was investigated at 22 °C, 30 °C, and 35 °C. SEM analysis confirmed a rough, porous morphology for both materials, while BET results showed a considerably higher specific surface area for CM II than for CM I. Accordingly, CM II generally exhibited better adsorption performance.

The Dubinin-Radushkevich and Dubinin-Astakhov isotherm models successfully described the adsorption process. The calculated adsorption energies indicate that CHP adsorption is predominantly physical, while the applicability of the Dubinin-Astakhov model suggests the importance of surface heterogeneity.

The adsorption capacity decreased with increasing temperature, indicating that CHP adsorption was more favorable at 22 °C. Although the obtained adsorption capacities were moderate, the results show that low-cost carbon materials derived from biowaste have potential for mitigating organophosphate contaminants in aquatic systems. Further optimization of precursor selection, carbonization conditions, porosity, and surface chemistry is needed to improve their adsorption efficiency and

practical applicability.

Acknowledgment

This work was partially supported by the Ministry for Science, Technological Development and Innovation of the Republic of Serbia (451-03-33/2026-03/200017 and 451-03-34/2026-03/200146 and Ministry of Defense Republic of Serbia (VA-TT/1/22-24).

References

[1] ANIĆIJEVIĆ, V.J., et al., Stability and neurotoxic impact of organophosphate pesticides in aqueous environments. Vojnotehnički glasnik, 2025. **73**(1): p. 264–281.

[2] ANIĆIJEVIĆ, V.J., et al., „Esters of Organophosphorus Acids-Toxicity, Application and Removal from the Environment ". Scientific Technical Review, 2019. **69**(3): p. 15–29.

[3] ZHU, X., et al., Visible-light-induced direct thiolation at α-C (sp3)–H of ethers with disulfides using acridine red as photocatalyst. Organic letters, 2016. **18**(7): p. 1546–1549.

[4] JAKUBOV, T.S. and D.E. Mainwaring, *Modified Dubinin–Radushkevich/Dubinin–Astakhov Adsorption Equations.* Journal of Colloid and Interface Science, 2002. **252**(2): p. 263–268.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SYNTHESIS OF TITANIUM BASED POWDERS

SREĆKO STOPIĆ, BERND FRIEDRICH
IME Process Metallurgy and Metal Recycling, RWTH Aachen University, Aachen,
ORCID: 0000-0002-1752-5378, sstopic@ime-aachen.de; bfriedrich@ime-aachen.de

DUŠKO KOSTIĆ; MITAR PERUŠIĆ,
Faculty of Technology Zvornik, University of the East Sarajevo, dusko.kostic@tgfzv.ues.rs.ba; mitar.perusic@tfzv.ues.rs.ba

VLADIMIR DAMJANOVIĆ, RADISLAV FILIPOVIĆ
Nova Alumina, Zvornik; vladimir.damjanovic@birac.ba, radislav.filipovic@birac.ba

MIRKO STAMATOVIĆ
RMD 18 Termoplastic, Uzice, stamatovic.mirko@gmail.com

DEJAN JOVANOVIĆ
Fero Promet, Užice, feropromet.uzice@gmail.com

Abstract: *Titanium is classified as a critical element and its application is very important in many industrial sectors. The synthesis of titanium compounds in this paper includes the synthesis of titanium oxide, ruthenium oxide-titanium oxide, core (titanium oxide)/shell(gold) particles and TiAlV-alloy. The synthesis is carried out using ultrasonic pyrolysis and aluminothermic reduction. As starting material, titanium oxy sulphate and tetra-n butylorthotitanat were used. Ultrasonic spray pyrolysis(USP) is simple technique for synthesizing nanoparticles and thin films,via ultrasonic atomization to create fine aerosol droplets that undergo thermal decomposition. Aluminothermic reduction represents an exothermic reeaction using aluminium as the reducing agent. Although the preparation of pure titanium is difficult because of its reactivity and high tendency to form oxides, some innovative solutions will be considered such as molten salt electrolysis. The great affinity of titanium towards oxygen is the biggest problem of how to obtain titanium metal powder and prevent oxidation.*

Keywords: *titanium, ultrasonic spray pyrolysis, synthesis*

1. INTRODUCTION

Titanium (Ti) is a transitional metal with an atomic number of 22 and is an especially interesting element due to its properties and the properties of the alloys it makes. Titanium is a low-density metal (4,5 g/cm^3) and possess high strength and corrosion and creep resistance, which makes it an attractive metal either in pure form or an alloy for various advance applications such as aero space engineering (light weight, high strength), energy production (as catalysts for solid oxide cells and various photo reactive coating) biomedical devices (high strength, corrosion resistance) and various other industry sectors [1].

The synthesis of metal and metal compounds particles plays a crucial role in applications such as protective coating, catalysis, electronic materials, energy production and storage. Among various producing techniques, Ultrasonic Spray Pyrolysis (USP) has gained attention due to its ability to produce homogeneous particles with controlled size, composition, continuous and low-cost production [2].

Thw main aim is to present synthesis of titanium oxide, titanium oxide-ruthenium oxide and gold titanium oxide powders using ultrasonic spray pyrolysis method. The second way is alumiothermic reduction for production of Titanium aluminium vanadium (TiAlV) alloys using reduction of titanium oxide and vanadium oxide.

2. EXPERIMENTAL

2.1. Synthesis of titanium oxide

Titanium oxysulfate ($TiOSO_4$); Venator, Duisburg, Germany solution was tested for the synthesis of nanosized titanium oxide using an ultrasonic generator (PRIZNANO, Kragujevac, Serbia) to produce aerosols, highlighting the scalability and versatility of the USP method for oxide nanopowder synthesis. Most important part of this experimental setup are: pump for an injection of precursor solution, ultrasonic atomizer, furnace with 4 heating zones, reaction tube, and an electrostatic precipitator, as shown at Figure 1.

Corresponding author: S. Stopić;
E-mail: sstopic@ime-aachen.de

DOI: 10.5937/OTEH26110S

Figure 1. Experimental setup for USP-synthesis with an electrostatic precipitator

The experimental preparation of titanium oxide powders from aqueous solutions of titanium oxysulfate, Venator, Duisburg was carried at a temperature of 1000°C in a compressed air atmosphere, with a solution flow rate of 5 l/min. The obtained particles are shown at Figure 2.

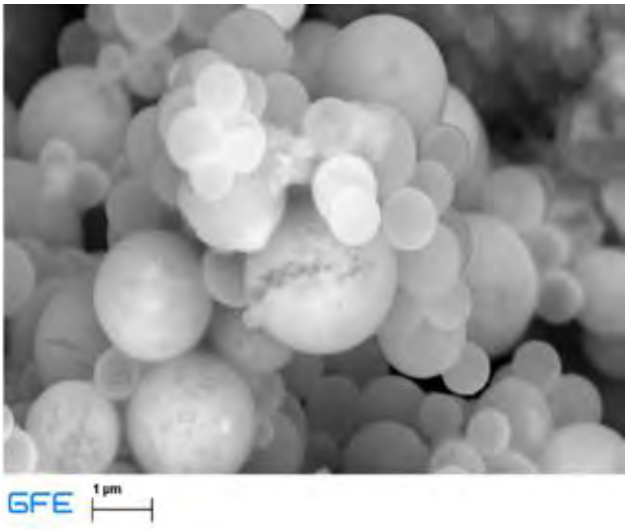


Figure 2. SEM Analysis of TiO_2-powders obtained from titanium oxysulphate by USP

The SEM image shows spherical particles formed through the ultrasonic spray pyrolysis process. The particles appear uniform in shape, with a size range in the micrometer scale, approximately 0.5–1 µm for ones, indicating possible agglomeration during or after formation [2]

2.2. Synthesis of Titanium oxide /ruthenium oxide powders

A new precursor was investigated in order to perpare titanium oxide. Ruthenium (III) chloride hydrate, hydrochloric acid, tetra-n-bytil orthotitanate and water were used as precursor for the synthesis of RuO_2/TiO_2 nanoparticles by USP using the equipment as shown in our previous publications [3]. The most important parts of the set up are the ultrasonic atomizer, the reactor and wet scrubber as the particle collector. The temperature and pressure control was adjusted using a thermostat and a vacuum pump. Atomization of the obtained solution after dissolution of precursor took place in an ultrasonic atomizer (Gapusol 9001, RBI/ France) in order to create the aerosol, as shown at Figure 3.

Figure 3. Experimental setup for USP-synthesis with wet scrubber

Regarding to our previous results the resonant frequency was selected between 0.8 MHz and 2.5 MHz. Under spray pyrolysis conditions a carrier gas flows continuously through the quartz tube at a certain flow rate. Atomized precursor droplets generated by the ultrasonic generator are further transported by carrier gas to the furnace. After thermal decomposition in the furnace, the Ru-Ti based nanoparticles were collected. The synthesis was performed in a nitrogen or oxygen atmosphere with different molar ratios of Ru and Ti: 1.0; 0.33; 0.22, and 0.17 in the precursor solution. The proposed mechanism of formation of catalyst carrier (or core material TiO_2) was shown on Fig. 4:

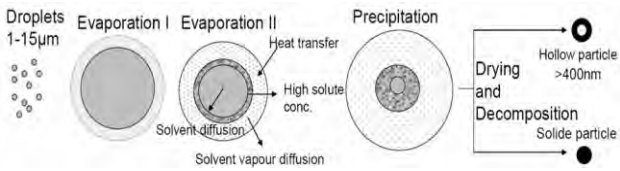


Figure 4: Formation mechanism for the synthesis of nanosized TiO_2 in single step

The chemism of TiO_2 formation is presented with Eq. 1:

$$0.5\ C_{16}H_{36}O_4Ti + HCl + H_2O = 0.5\ TiO_2 + 2\ C_4H_9OH + 0.5\ Cl_2 \quad (1)$$

During evaporation of one aerosol droplet, following precipitation from solution, oxidation and sintering the different form and size can be formed depending on the used reaction parameters. Model of one droplet to one nanoparticle was accepted in thid synthesis. The RuO_2/TiO_2 nanoparticle synthesis was performed at different operating parameters. The chemical composition of of RuO_2/TiO_2 powders with structural analysis is shown in Table 1.

Table 1: Rietveld XRD analysis of obtained powder

Molar ratio of Ru and Ti in the solution	Anatase TiO_2 (%)	Rutile TiO_2 (%)	RuO_2
0.33	5.1	63.8	31.1
0.22	-	83.7	16.3
0.17	-	87.8	12,2

Fig. 5 displays SEM images from a FIB section of RuO_2/TiO_2 nanospheres obtained at 800°C. The dark areas indicate the TiO_2 core, which consists of several sub-particles, each of which is decorated by ruthenium-rich particles of the size of 1-5 nm.

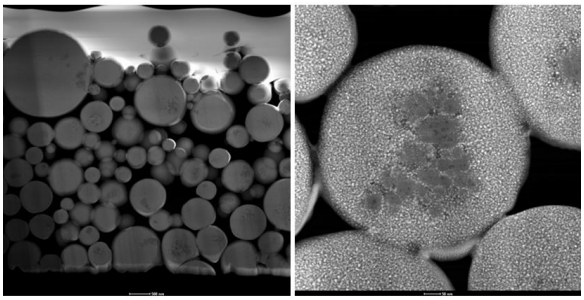


Figure 5: FIB and SEM analysis of obtained powder (T=800°C, ratio 0.333; f=2.5 MHz)

STEM/EDS analysis of the dark areas confirmed a high titanium concentration, but also the presence of Ru. Separated areas containing pure titanium oxide or ruthenium oxide were not detected. It confirms that is very difficult to obtain ideally core (TiO_2)/shell (RuO_2) structure using ultrasonic spray pyrolysis method.

2.3. Synthesis of gold/titanium oxide using ultrasonic spray pyrolysis

Synthesis of gold/titanium oxide particles was performed using hydrogen tetrachloroaurate ($HAuCl_4$) with Tetra-n butylorthotitanat. Very fine particles were prepared in hydrogen and nitrogen atmosphere using same equipment shown at Figure 3, The obtained core/shell particles ere shown at Figure 6 with their particle size distribution on Figure 7.

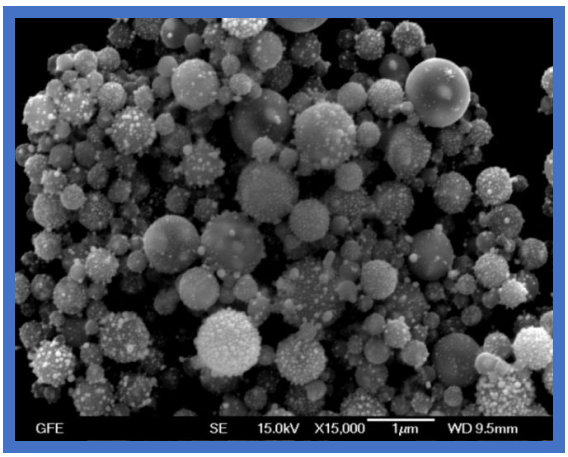


Figure 6: Au/TiO_2 particles obtained by USP

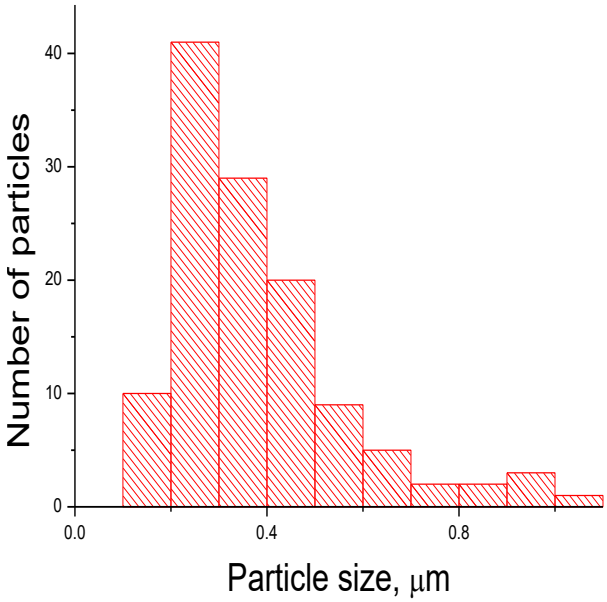


Figure 7. Particle size distribution of Au/TiO_2 particles

The obtained particles contain core (TiO_2) and shell (Au).. Average diameter of core (TiO_2) amounts about 350 nm, and smaller than 100 nm for shell (Au). G. Alkan et al [4] explains the mechanism of the metal/oxide core-shell Ag/TiO_2 and Au/TiO_2 nanoparticle formation via one-step ultrasonic spray pyrolysis (USP) by establishing a new model. The general knowledge on the standard "droplet-to-particle" (DTP) mechanism, nucleation, and growth processes of noble metals, as well as physical and chemical properties of core and shell materials and experimental knowledge, were utilized with the purpose of the construction of this new model.

2.4. Synthesis of TiAlV alloys from TiO_2 and V_2O_5 using aluminothermic reduction

Aluminothermic reduction is based on self-propagating high temperature reaction through reduction reaction of titanium oxide and vanadium oxide using aluminum powder. The reduction of titanium oxide with aluminum does not provide enough heat for self-propagation. The heat released from the reaction mixture must therefore be regulated by adding boosters such as $KClO_4$. Stoichiometrical values of Al to TiO_2 and V_2O_5 varied in order to prepare Ti60Al24V16 alloys. In order to improve the reaction behaviour and the separation of slag and metal, calcium oxide was added to the reaction vessel to adjust the melting point of the resulting slag system. The experimental setup is shown in Figure 8:

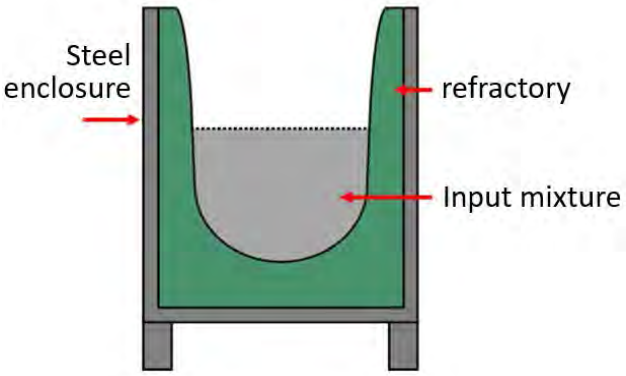


Figure 8: Experimental setup for aluminothermic reduction.

In contrast to ideal spherical particles of TiO_2 obtained by ultrasonic spray pyrolysis. the particles of TiAlV-powder have irregular form as shown at Figure 9.

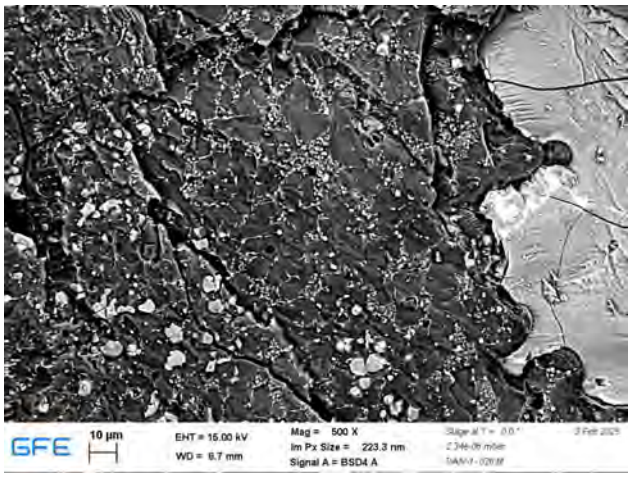


Figure 9: SEM Analysis of TiAlV-alloy obtained using aluminothermic reduction

Cheng et al. [5] mentioned the relevant equations in accordance with stoichiometry for synthesizing Ti-6Al-4V alloy are as follows (the molar ratio of CaO/Al_2O_3 is 0.4), and the calculated adiabatic temperature amounts 1750 K. Mechanism of process is presented with equations (2)-(5).

$$3\ TiO_2 + 4\ Al \rightarrow 3\ Ti + 2\ Al_2O_3 \quad (2)$$

$$3\ V_2O_5 + 10\ Al \rightarrow 6\ V + 5\ Al_2O_3 \quad (3)$$

$$TiO_2 + 0.021\ V_2O_5 + 1.52\ Al \rightarrow 3\ Ti\text{-}6Al\text{-}4V + 0.70\ Al_2O_3 \quad (4)$$

$$TiO_2 + 0.021\ V_2O_5 + 1.52\ Al + 0.28CaO \rightarrow 3Ti\text{-}6Al\text{-}4V + 0.25\ CaO{\cdot}2Al_2O_3 + 0.035CaO{\cdot}6Al_2O_3 \quad (5)$$

3. CONCLUSION

The synthesis of titanium oxide, ruthenium oxide/titanium oxide and gold/titanium oxide powders was successfully performed using the ultrasonic pyrolysis process from different precursors. Precursors such as titanium oxysulfate and tetra-n-bytil orthotitanate have been successfully chosen for the formation of titanium-based compounds by USP. Due to the great affinity of titanium towards oxygen, the reduction of titanium oxide and vanadium oxide was carried out using aluminum powder, producing an alloy of aluminum, titanium and vanadium.

References

[1] LÜTJERING, G, WILLIAMS, J.C, Titanium, 2nd ed. Berlin, Germany: Springer, 2007.

[2] STOPIC, S., KOSTIC, D., DAMJANOVIC, V, PERUŠIC, M., FILIPOVIC, R., NIKOLIC, N, FRIEDRICH, B. Recovery of Titanium and Aluminum from Secondary Waste Solutions via Ultrasonic Spray Pyrolysis, Metals 2025, 15, 701 https://doi.org/10.3390/met15070701

[3] KOŠEVIĆ, M., STOPIC, S., BULAN, A., KINTRUP, J., WEBER, R., STEVANOVIĆ, J., PANIC, V., FRIEDRICH, B (2017): A continuous process for the ultrasonic spray pyrolysis synthesis of RuO_2-TiO_2 particles and their application as an active coating of activated titanium anode, Advanced Powder Technology, 28, 43-49.

[4] ALKAN, G., RUDOLF; R, BOGOVIC; JENKO, D., FRIEDRICH, B, Ultrasonic Spray Pyrolysis Structure and Formation Model of Ag/TiO2 and Au/TiO2 Nanoparticles Synthesized through Ultrasonic Spray Pyrolysis, Metals 2017, 7, 389;

[5] CHENG, C., DOU, Z. H., ZHANG, T.A., ZHANG, H. J, YI, X., SU, J. M (2017): Synthesis of As-cast Ti-Al-V alloy from titanium-rich material by thermite reduction, *JOM*, 1-6. DOI: 10.1007/s11837-017-2467-7

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 7 - 10 October 2026

POROUS CERAMIC FOAM AS A MEDIUM FOR COBALT IMMOBILIZATION IN RADIONUCLIDE WATER TREATMENT

DENIS DINIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, denis.dinic@yahoo.com

MILAN TANIĆ
The Academy of Applied Preschool Teaching and Health Studies, Department of Technical and Technological Studies Kruševac, Kruševac, milantanic@yahoo.com

JELENA DIKIĆ
Innovation Centre of Faculty of Technology and Metallurgy, Belgrade, jdikic@tmf.bg.ac.rs

STEVAN STUPAR
University of Criminal Investigation and Police Studies, Belgrade, stevan.stupar13@gmail.com

NEMANJA JOVANOVIĆ
Technical Faculty in Bor, University of Belgrade, Bor, nemanjanikjov77@gmail.com

SANJA JEVTIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, sanja@tmf.bg.ac.rs

Abstract: *Radionuclides in the environment are a significant concern, especially due to the increasing release of radioactive cobalt (^{60}Co) into water streams in emergencies. This trend presents considerable environmental risks that require urgent attention. There is growing interest in the development of low-cost inorganic ion-exchange materials for the treatment of various types of radioactive liquid waste. This study investigates an efficient method for the direct immobilization of cobalt (Co^{2+}) ions onto porous ceramic foam based on zeolite and bentonite. The aim of the present research was to establish a simplified approach for water purification and remediation in emergencies. The material was synthesized using an eco-friendly foam method and subsequently sintered at 800°C, a temperature significantly higher than that typically employed in conventional calcination methods for zeolites. The characteristics of the prepared porous material were analyzed using scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDX). The adsorption behavior of Co^{2+} ions onto the porous ceramic material was found to vary significantly in the presence of surfactants. To assess the leachability of the adsorbed Co^{2+} ions, a simulated acid rain test procedure was employed. Consequently, the porous ceramic foam, based on natural zeolite, provides a viable means for processing radioactive waste in an environmentally friendly manner.*

Keywords: *zeolite, bentonite, adsorption, remediation, surfactant.*

1. INTRODUCTION

To reduce global warming emissions, many countries are shifting from fossil fuels toward nuclear and renewable energy. Nuclear power offers high energy output with minimal greenhouse gas emissions. However, its major disadvantage is the potential generation of hazardous radioactive materials in emergencies during accidents, malfunctions, or decommissioning, posing risks to human health and the environment [1].

Radioisotopes realised in accidents can easily be dissolved in water [2,3], which can lead to rapid environmental contamination.

In addition, radioactive wastewater is often produced during the decontamination of contaminated surfaces. Such radioactive contaminated water typically contains cesium, strontium, and cobalt, as well as suspended organic matter originating from surfactants used in the decontamination process [4].

Cobalt is of particular concern, as it is a highly toxic radionuclide, and the removal of Co^{2+} from radioactive water has received increasing attention in recent years [5].

Regarding to the remediation of aqueous systems contaminated with radionuclides, radioactive water must be properly treated and converted into a stable solid form prior to disposal.

Remediation of radionuclide contaminated water increasingly prioritizes approaches that are technically effective, sustainable, and low-maintenance. According to IAEA guidance (IAEA-475) [6], remediation strategies should be selected based on thorough site characterization and, where feasible, the use of passive treatment methods.

Corresponding author: D. Dinić;
E-mail: denis.dinic@yahoo.com

DOI: 10.5937/OTEH26111D

In recent years, extensive research has focused on inorganic ion-exchange materials capable of adsorbing radionuclides from wastewater. Conventional organic resins exhibit several limitations, including poor heat resistance, insufficient radiation stability, and difficulties associated with the treatment of secondary solid waste [5,7].

In contrast, inorganic ion-exchange materials offer several advantages: lower cost, strong resistance to radiation damage, higher ion-exchange and adsorption capacities, and the ability to directly solidify secondary waste through grouting [2]. Although organic compounds such as surfactants can influence the adsorption behavior [7,8].

Porous ceramic materials, which rank among the most commonly used inorganic materials today, have shown great potential as adsorbents for various metal ions [9]. In general, porous ceramics are defined as materials with porosity levels exceeding 30% [10]. Within this category, two primary ceramic structural types are distinguished: 2D honeycomb and 3D foam structures. A variety of fabrication techniques have been developed for producing ceramic foams, including the replica template method, sacrificial template method, and direct foaming [11].

This study investigates the effects of suspended organics and Co^{2+} ions on the adsorption performance of ceramic foam, with the aim of improving their application in radioactive water remediation.

2. EXPERIMENTAL

2.1. Materials and chemicals

The materials used for the synthesis of porous ceramics were natural zeolite from the Slanci deposit (Serbia), natural bentonite with a particle size of 63 µm provided by Bentoprodukt d.o.o., Šipovo (Bosnia and Herzegovina), sodium dodecyl sulfate (SDS) as the foaming agent (Merck, Germany), a hydrogen peroxide solution (H_2O_2, 30%, Centrohem, Serbia) as the blowing agent. Analytical grade salts cobalt(II) nitrate hexahydrate ($Co(NO_3)_2 \cdot 6H_2O$, ≥99.0%) was purchased from Merck (Germany). Sulfuric acid (H_2SO_4, 95.0-97.0%), nitric acid (HNO_3, 65.0%), and sodium chloride (NaCl) were obtained from Centrohem (Serbia). Three types of surfactants were used: SAFOL 23E6.5 (SAFOL) consisted of C_{12}–C_{13} oxo alcohol + 6.5 mol ethylene oxide/mol (Sasol Chemical Industries, South Africa) as non-ionic surfactants, cetyltrimethylammonium chloride (HDTMA, Acros Organics, USA) as cationic surfactants, and sodium dodecyl sulfate (SDS, Merck, Germany) as anionic surfactants.

2.2. Sample preparation

The porous ceramics samples were prepared according to the procedure described in our previous study [12]. The samples were prepared using the foam method. The resulting samples of porous ceramics made from natural zeolite and bentonite clay after sintering with an average porosity of 63%, are referred to as CF.

2.1. Batch adsorption and desorption experiments

The effects of surfactants on the adsorption of Co^{2+} was investigated. In all experiments, 1 g CF was added into 100 mL simulated radioactive solution containing Co^{2+} ions and different surfactants. The initial concentrations of Co^{2+} ions were applied at 100 mg L^{-1}.

Surfactants SAFOL 23 E6.5, HDTMA, and SDS were introduced to Co^{2+} ions solution at their respective critical micelle concentrations 0.037 mmol L^{-1}, 1 mmol L^{-1}, and 8.2 mmol L^{-1}, respectively. After mixing 1 g of CF with the respective solutions for 24 h, the solutions were filtered through a 0.45 µm PTFE syringe filter. The filtrates were subsequently diluted with deionized water prior to AAS analysis. Adsorption capacity and removal efficiency were calculated using Eq. (1) and Eq. (2):

$$Q_e = \frac{(C_0 - C_e)V}{m} \quad (1)$$

$$(\%) = \frac{C_0 - C_e}{C_0} x100 \quad (2)$$

where Q_e (mg g^{-1}) is the adsorption capacity; C_e (mg L^{-1}) is the concentration in the equilibrium solution; C_0 (mg L^{-1}) is the initial concentration of the solution, V (L) is the volume of the added solution, and m (g) is the mass of adsorbent.

Further, CF saturated with Co^{2+} and surfactant was shaken in a 100 mg L^{-1} simulated acid rain (SAR) solution for 24 h for desorption testing. The SAR solution was prepared using a mixture of H_2SO_4 and HNO_3 adjusted to pH 5.6 [13]. Desorption experiments were conducted at 105 rpm and 25 °C for 24 h. The concentration of metal ions in the supernatant was measured by AAS, and the desorption rate was calculated using Eq. (3).

$$W = \frac{m_d}{m_a} x100\% \quad (3)$$

where W (%) is the desorption rate; m_d (g) is the amount of metal ions in the desorption solution; m_a (g) is the amount of adsorbed metal ions by the sintered samples.

2.4. Analytical method

The morphology of the sintered samples was examined by scanning electron microscopy (SEM) with Energy-Dispersive X-ray (EDX) Spectroscopy (JEOL 6610LV, Japan). The concentration of metal ions in aqueous solutions during adsorption and desorption experiments was measured by an atomic absorption spectrometer (AAS, SpectrAA 55B, Varian, USA). A Vanta Max series portable handheld X-ray fluorescence (XRF) spectrometer (Olympus, USA) evaluated the metal ions adsorbed onto sintered samples.

3. RESULTS AND DISCUSSION

The morphology of the sintered sample CF is shown in Fig. 1. The micrographs reveal a heterogeneous surface characterized by pores of varying sizes distributed throughout the structure. The overall morphology appears relatively rough, indicating an irregular sintering profile. The EDX spectrum exhibits pronounced peaks corresponding to Si, Al, and Fe, confirming the presence of silicon-, aluminum-, and iron-bearing oxides within the sample. Notably, no elevated carbon signal is detected, supporting the conclusion that organic constituents were fully decomposed during the sintering process. The EDX analysis of is shown in Table 1 to indicate the atomic and oxide percentage.

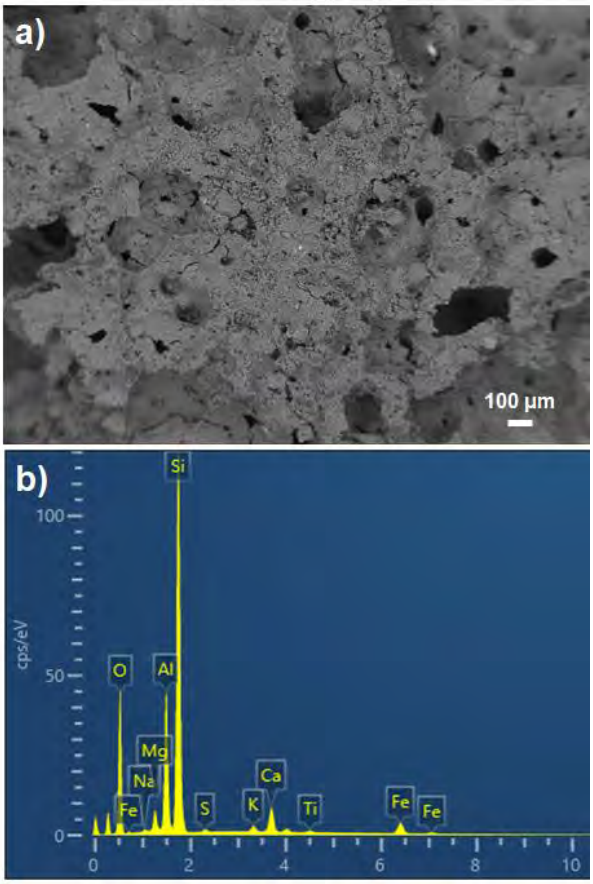


Figure 1. SEM and EDX images of the CF.

Table 1. EDX analysis of the CF.

Element	Atomic (%)	Oxide	Oxide (%)
O	63.52		
Si	22.8	SiO_2	65.37
Al	8.1	Al_2O_3	19.71
Ca	1.75	CaO	4.68
Fe	1.49	FeO	5.12
Mg	1.46	MgO	2.8
K	0.36	K_2O	0.81
Na	0.22	Na_2O	0.33
S	0.17	SO_3	0.65
Total	100	Total	100

In this study, stable Co^{2+} ions replaced radioactive ^{60}Co to investigate cobalt adsorption behavior in aqueous systems. The use of stable Co^{2+} provides the same relevant chemical properties and adsorption characteristics as the ionic form of cobalt found in radioactive wastewater, while avoiding the radiological hazards and additional safety requirements associated with handling ^{60}Co. Therefore, stable Co^{2+} ions were selected as a safe and practical substitute for studying cobalt adsorption under controlled laboratory conditions. However, it should be noted that the experiments conducted do not account for the radiological characteristics of ^{60}Co, such as activity and radioactive decay.

Previous studies have demonstrated that ceramic materials are effective adsorbents for the treatment of radioactive wastewater [8]. However, the influence of surfactants commonly present in such wastewater on the adsorption behavior of metal ions onto ceramic foam has not been comprehensively examined. In incident crisis, detergents used to decontaminate workers' bodies and protective clothing contain various surfactants that may interact with both radionuclide ions and the ceramic surface. This study therefore examines how various surfactant types influence Co^{2+} ions adsorption onto CF (Fig. 2).

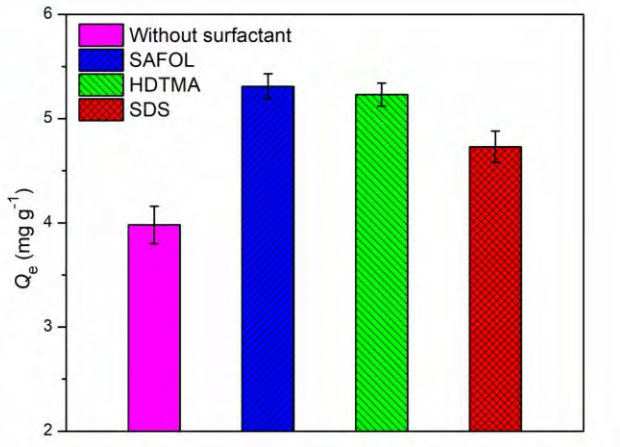


Figure 2. The effect of surfactants on the adsorption capacity values of Co^{2+}.

The presence of surfactants highly affected the adsorption of Co^{2+} ions onto CF. When surfactants were present in the metal solutions, the adsorption of Co^{2+} sharply increased. For non-ionic surfactants, SAFOL had the most significant effect on the adsorption of Co^{2+} onto ceramic foam, compared to other surfactants. The equilibrium adsorption capacities of SAFOL, HDTMA and SDS after 24 h of contact were 5.31 mg g^{-1}, 5.23 mg g^{-1} and 4.73 mg g^{-1}, respectively. The higher adsorption observed with SAFOL may be attributed to its non-ionic character, which can modify the CF surface and facilitate the interaction of Co^{2+} ions with available adsorption sites. The formation of a surfactant layer may also provide additional favorable sites for Co^{2+} uptake.

Parka et al. [8] investigated the efficiency of ammonium molybdophosphate–polyacrylonitrile composites for the adsorption of cobalt, strontium, and cesium ions from radioactive laundry wastewater. Their findings demonstrated that the introduction of nonionic, cationic, and anionic surfactants–specifically Triton-X, HDTMA, and SOBS–at concentrations corresponding to their CMC

values significantly enhanced the adsorption of Co^{2+} ions. This improvement was attributed to surfactant-induced modifications of the adsorbent surface and altered interactions between the adsorbent matrix and metal ions, highlighting the potential of surfactant-assisted adsorption as an effective strategy for radionuclide removal. The present results are consistent with these findings, while the slightly higher adsorption observed for SAFOL suggests that its non-ionic nature may provide more favorable conditions for Co^{2+} adsorption onto CF.

The XRF spectra of the sintered samples before and after adsorption are shown in Fig. 3.

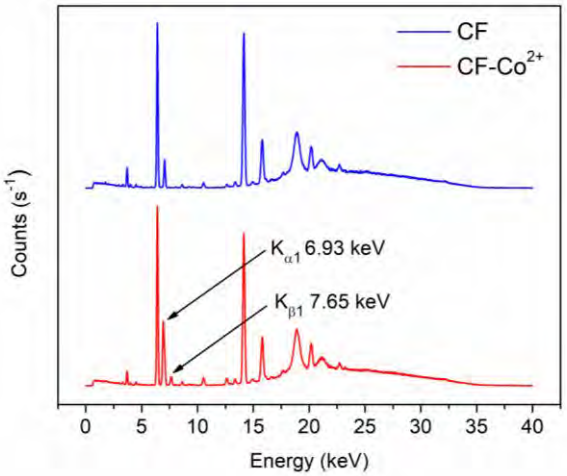


Figure 3. The XRF spectra of the sintered samples before and after adsorption of Co ions.

The XRF spectra of the sintered samples before and after adsorption. The XRF spectra shows that the peaks are moved when spectra's of untreated ceramic foam and sample after the adsorption of Co^{2+} ions, without interference of surfactants, are compared. Characteristic peaks of cobalt can be observed in sintered sample which is according by available literature [14]. The detection of cobalt directly on the CF surface by XRF provides additional evidence that Co^{2+} ions were effectively retained by the ceramic foam during the adsorption process.

The amount of cobalt in CF was evaluated using XRF to confirm and verify the existence of the maximum experimentally observed Qe. The XRF results presented in Fig. 4 showed a good correlation with the AAS results, confirming the change in metal ion concentrations in the aqueous solutions after adsorption from the solution of 100 mg L^{-1} concentration at 25 °C after 24 h.

This consistency between the XRF and AAS results supports the reliability of the experimentally determined adsorption capacity and confirms the occurrence of Co^{2+} uptake by CF.

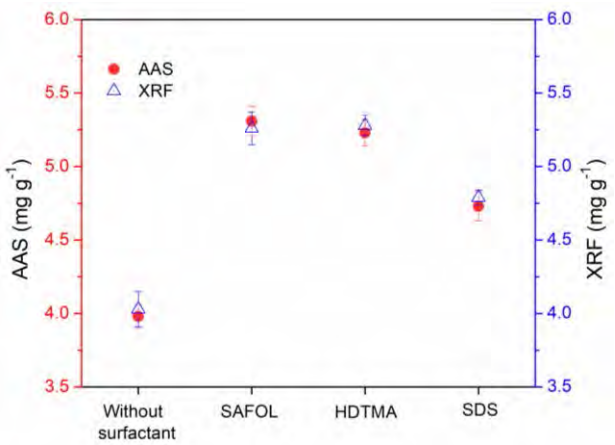


Figure 4. Comparison between concentration measured by AAS and fluorescence intensity measured by XRF for heavy metal ions adsorbed onto sintered samples.

In the SAR solution, no desorption of Co^{2+} ions from the ceramic foam was observed, indicating strong binding between the ions and the ceramic surface. The Co^{2+} concentration in the SAR solution was below the detection limit of the AAS method (0.05 mg L^{-1}). These findings suggest that the adsorbed samples can be considered suitable materials for the immobilization of Co^{2+} ions and may offer an environmentally safe option for their long-term containment.

5. CONCLUSION

A ceramic foam derived from natural zeolite and bentonite was synthesized and evaluated for its efficiency in the removal of cobalt (Co^{2+}) ions from aqueous solutions, simulating radionuclide-contaminated water.

The study confirmed the applicability of ceramic foam for removing Co^{2+} from simulated radioactive water. In all cases, the presence of surfactants enhanced Co^{2+} adsorption onto the ceramic foam, with similar overall adsorption trends observed across the three surfactant types.

The most pronounced improvement occurred with the non-ionic surfactant, which yielded the highest adsorption capacity of 5.31 mg g^{-1}.

In conclusion, the adsorption characteristics of the zeolite–bentonite ceramic foam demonstrate qualities that correspond to immobilization of Co^{2+} ions and sorption-based remediation strategies.

References

The example is given below.

[1] PULPEA, D., ROTARIU, T., TOADER, G., PULPEA, G.B., NECULAE, V., TEODORESCU, M.: *Decontamination of radioactive hazardous materials by using novel biodegradable strippable coatings and new generation complexing agents*, Chemosphere, 258 (2020) 127227.

[2] DE HARO-DEL RIO, D.A., AL-JOUBORI, S., KONTOGIANNIS, O., PAPADATOS-GIGANTES, D., AJAYI, O., LI, C., HOLMES,

S.M.: *The removal of caesium ions using supported clinoptilolite*, Journal of Hazardous Materials, 289 (2015) 1-8.

[3] NAGASAKI, S., NAKAYAMA, S.: *Radioactive Waste Engineering and Management*, Springer Japan, Tokyo, 2015.

[4] LIU, S., HE, Y., XIE, H., GE, Y., LIN, Y., YAO, Z., JIN, M., LIU, J., CHEN, X., SUN, Y., WANG, B.: *A State-of-the-Art Review of Radioactive Decontamination Technologies: Facing the Upcoming Wave of Decommissioning and Dismantling of Nuclear Facilities*, Sustainability, 14 (7) (2022) 4021.

[5] ZHU, Y., HU, J., WANG, J.: *Removal of Co2+ from radioactive wastewater by polyvinyl alcohol (PVA)/chitosan magnetic composite*, Progress in Nuclear Energy, 71 (2014) 172/178.

[6] The International Atomic Energy Agency, IAEA-475: Guidelines for Remediation Strategies to Reduce the Radiological Consequences of Environmental Contamination (IAEA Technical Reports Series No. 475, 2012).

[7] FANG, X., XU, Z., LUO, Y., REN, L., HUA, W.: *Removal of Radionuclides from Laundry Wastewater Containing Organics and Suspended Solids Using Inorganic Ion Exchanger*, Procedia Environmental Sciences, 31 (2016) 375-381.

[8] PARK, Y., LEE, Y.-C., SHIN, W.S., CHOI, S.-J.: *Removal of cobalt, strontium and cesium from radioactive laundry wastewater by ammonium molybdophosphate–polyacrylonitrile (AMP–PAN)*, Chemical Engineering Journal, 162 (2010) 685-695.

[9] SAREMI, O., GHAANI, M.R., KESHAVARZ, L., ENGLISH, N.J.: *Application of Porous Ceramics*, in: Advanced Functional Porous Materials, Springer, (2022) 499-537.

[10] OBADA, D.O., DODOO-ARHIN, D., DAUDA, M., ANAFI, F.O., AHMED, A.S., AJAYI, O.A.: *Potentials of fabricating porous ceramic bodies from kaolin for catalytic substrate applications*, Applied Clay Science, 132-133 (2016) 194-204.

[11] STUDART, A.R., GONZENBACH, U.T., TERVOORT, E., GAUCKLER, L.J.: *Processing Routes to Macroporous Ceramics: A Review*, Journal of the American Ceramic Society, 89 (2006) 1771-1789.

[12] DINIC, D., SOKIC, K., DIKIC, J., NIKOLIC, V., MARINKOVIC, A., ANDJIC, Z., KAMBEROVIC, Z., JEVTIC, S.: *Porous ceramic based on cost-effective natural zeolite and kaolin for removal of ammonium ions: Preparation, characterization and application,* Science of Sintering, (2024) 38-38.

[13] Zhao, C., Ren, S., Zuo, Q., Wang, S., Zhou, Y., Liu, W., Liang, S.: *Effect of nanohydroxyapatite on cadmium leaching and environmental risks under simulated acid rain*, Science of the Total Environment, 627 (2018) 553-560.

[14] BECKHOFF, HABIL. B., KANNGIEßER, B., LANGHOFF, N., WEDELL, R., WOLFF, H.: *Handbook of Practical X-Ray Fluorescence Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF SURFACE PREPARATION ON THE MECHANICAL PROPERTIES OF ADHESIVELY BONDED JOINTS IN METAL STRUCTURES

NATAŠA ZDRAVKOVIĆ
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia, natasa.zdravkovic@masfak.ni.ac.rs

DRAGAN MILČIĆ
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia, dragan.milcic@masfak.ni.ac.rs

DAMJAN KLOBČAR
Faculty of Mechanical Engineering, University of Ljubljana, Ljubljana, Slovenia, damjan.klobcar@fs.uni-lj.si

TOMAŽ VUHERER
Faculty of Mechanical Engineering, University of Maribor, Maribor, Slovenia, tomaz.vuherer@um.si

MIODRAG MILČIĆ
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia, miodrag.milcic@masfak.ni.ac.rs

VUKAŠIN PAVLOVIĆ
Faculty of Mechanical Engineering, University of Niš, Niš, Serbia, vukasin.pavlovic@masfak.ni.ac.rs

ALEKSIJA ĐURIĆ
Faculty of Mechanical Engineering, University of East Sarajevo, East Sarajevo, Bosnia and Herzegovina, aleksija.djuric@ues.rs.ba

Abstract: *Adhesively bonded joints represent a reliable method of joining materials in modern structures, and the development of new adhesives has enabled their application under demanding service conditions, including defense systems and military equipment. This study examines the effects of manual sanding, laser cleaning, and laser surface texturing on the lap shear strength of single-lap adhesively bonded joints fabricated from EN AW 5754 aluminum alloy and X5CrNi18-10 stainless steel using three structural adhesives. The results showed that laser cleaning provided the highest lap shear strength, while conventional sanding proved optimal for all adhesives investigated. Joints made of stainless steel exhibited 16% to 40% higher strength compared to joints made of aluminum alloy under identical surface preparation conditions using the same adhesive. The results contribute to a better understanding of the relationship between laser surface treatment and the mechanical performance of adhesively bonded joints and provide guidance for their application in defense-related structures, where high requirements are imposed on reliability, durability, and resistance to dynamic loading.*

Keywords: *adhesive bonded joints, surface preparation, lap shear strength, aluminum alloys, stainless steel.*

1. INTRODUCTION

The requirements of the modern defense industry for lightweight, reliable, and high-performance structural joints have driven intensive development in the field of adhesive metal bonding. Defense platforms, ranging from armored vehicles and aircraft to naval vessels and ballistic protection systems, increasingly employ multi-material structures in which bonded joints offer advantages that traditional mechanical fasteners or welding cannot provide: a more uniform stress distribution, reduced structural mass, excellent fatigue resistance, and the elimination of stress concentrations in the joint zone [1, 2, 3].

Surface preparation prior to bonding is a key factor determining the quality and durability of adhesive joints. In addition to conventional mechanical methods, such as sanding and sandblasting, and chemical etching, laser surface treatment has emerged as an environmentally friendly, repeatable, and highly controllable alternative [4]. Depending on the selected processing parameters, laser cleaning and laser texturing can increase surface roughness, enhance surface wettability by inducing superhydrophilic behavior, and modify the chemical composition of the surface layer, thereby contributing to improved joint strength [5, 6].

The materials most commonly used in defense structures, namely 5000-series aluminum alloys and austenitic stainless steels, present particular challenges for adhesive bonding. Aluminum alloys are characterized by the

Corresponding author: N. Zdravković;
E-mail: natasa.zdravkovic@masfak.ni.ac.rs

DOI: 10.5937/OTEH26112Z

presence of a stable oxide layer that adversely affects adhesive wetting, while stainless steels are difficult to bond without appropriate mechanical or chemical surface preparation [7]. The aim of this study was to investigate the influence of different mechanical and laser surface preparation methods on the lap shear strength, wettability, and fracture mechanisms of adhesively bonded joints made from these materials, as well as to discuss their potential application in defense systems.

2. MATERIALS AND METHODS

2.1. Base materials and adhesives

The investigation was carried out on EN AW 5754 aluminum alloy (H22 temper) and X5CrNi18-10 stainless steel (EN 1.4301) plates, both 2 mm thick. Both materials were water-jet cut into specimens measuring 100 mm × 25 mm in accordance with EN 1465 [8]. Three structural adhesives were used for bonding: two two-component adhesives, SikaPower®-880 (2K, high fatigue resistance) and SikaFast®-580 (2K, acrylic-epoxy hybrid), and one one-component adhesive, SikaPower®-492 G (1K, high strength). The main characteristics of the investigated adhesives are summarized in Table 1. Single-lap joints with an overlap length of 12.5 mm were fabricated, and the adhesive layer thickness was maintained at 0.3 mm using spacers [9].

Table 1 Main characteristics of the investigated structural adhesives

Adhesive	Type	Chemical basis	Main characteristic
SikaPower®-880	2K	Epoxy adhesive	High fatigue resistance and structural bonding.
SikaFast®-580	2K	Acrylic-epoxy	Fast curing, high strength and impact resistance.
SikaPower®-492 G	1K	Heat-curing epoxy hybrid	High static strength for structural applications.

2.2 Surface preparation methods

Four surface preparation methods were applied: (1) manual abrasion using a Scotch-Brite™ abrasive pad (SB), (2) manual sanding with P180 abrasive paper (P180), (3) laser cleaning (LC) (Jetlaser M100, 45 W laser power, pulse energy of 1.25 mJ), and (4) laser texturing (LT) using a nanosecond pulsed fiber laser (FL-mark-C equipped with a JPT MOPA source, λ = 1064 nm, 30 W). Circular and linear surface patterns with depths of 40 µm and 80 µm were produced during the texturing process. Before the mechanical and laser texturing treatments, the specimens were degreased with SIKA® Remover-208, a solvent-based cleaning agent for heavily soiled, non-porous surfaces, intended for cleaning and degreasing surfaces prior to bonding.

2.3 Testing methods

The testing program included:

– surface roughness measurements using a Mitutoyo SJ-301 contact profilometer (Ra and Rz parameters);

– contact angle measurements using an Osilla goniometer with double-distilled water; and

– static lap shear strength testing of single-lap joints using an Instron 8802 universal testing machine (250 kN capacity) at a crosshead speed of 1 mm/min, in accordance with ISO 4587 [10].

Fracture surface analysis was performed using a high-resolution digital camera and the images were processed using ImageJ image processing software.

3. RESULTS AND DISCUSSION

3.1 Surface roughness and wettability

Surface roughness measurements were performed in three separate measurement regions for each specimen, in two mutually perpendicular directions: longitudinally and tangentially. Sanding with P180 abrasive paper resulted in higher surface roughness than Scotch-Brite™ preparation for both materials investigated. The average roughness values obtained after P180 sanding were Ra = 1.75 µm and Rz = 10.8 µm for the EN AW 5754 aluminum alloy, and Ra = 0.45 µm and Rz = 3.75 µm for X5CrNi18-10 stainless steel.

Laser texturing produced a superhydrophilic surface immediately after treatment, as evidenced by a contact angle of approximately 0°. However, the contact angle gradually increased during exposure to ambient air, which has been attributed to the adsorption of organosilicon compounds from the environment [11, 12].

3.2 Lap shear strength of adhesively bonded joints

The lap shear strength results for adhesively bonded joints prepared using different surface treatment methods are presented separately for the EN AW 5754 aluminum alloy (Figure 1) and X5CrNi18-10 stainless steel (Figure 2).

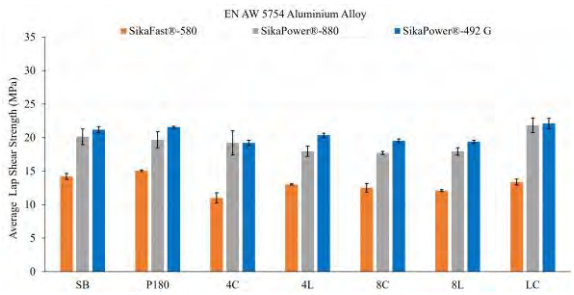


Figure 1. Average lap shear strength of single-lap adhesive joints prepared using different surface preparation methods for EN AW 5754 aluminum alloy (SB – Scotch-Brite; P180 – abrasive paper; 4C and 8C – circular patterns at 40 and 80 µm depth; 4L and 8L – linear patterns at 40 and 80 µm depth; LC – laser cleaning)

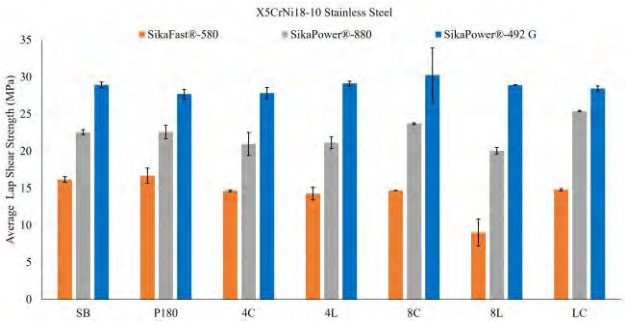


Figure 2. Average lap shear strength of single-lap adhesive joints prepared using different surface preparation methods for X5CrNi18-10 stainless steel

Laser cleaning provided the highest lap shear strength for the SikaPower®-880 adhesive, particularly for joints made of X5CrNi18-10 stainless steel, while the corresponding values for EN AW 5754 aluminum alloy were also among the highest obtained. For the SikaFast®-580 adhesive, sanding with P180 abrasive paper resulted in the highest shear strength. In the case of SikaPower®-492 G, P180 preparation provided high joint strength, although comparable or higher values were achieved after laser surface texturing, particularly for stainless steel.

For all three adhesives, joints made of X5CrNi18-10 stainless steel generally exhibited approximately 16–40% higher lap shear strength than those made of EN AW 5754 aluminum alloy under identical surface preparation conditions. This behavior may be attributed to differences in the surface characteristics of the two materials and their interaction with the adhesive [9].

3.3 Fracture mechanisms

Cohesive failure, which is generally considered an indicator of a high-quality adhesive joint, was the dominant failure mode for the SikaPower®-880 adhesive after laser cleaning and P180 sanding for both materials. In contrast, predominantly adhesive (interfacial) failure was observed for laser textured specimens, indicating insufficient adhesion. This behavior was associated with surface oxidation and the redeposition of molten particles generated during laser processing, as confirmed by the surface analyses performed in this study [9, 13, 14].

From the perspective of defense applications, the predominance of cohesive failure is particularly important for the reliability of adhesively bonded joints subjected to vibration and impact loading, such as ballistic impact and blast loading. In addition, laser surface preparation offers practical advantages for defense manufacturing due to its high repeatability, ease of automation and elimination of hazardous chemical cleaning agents.

4. CONCLUSION

Based on the experimental investigations, the following conclusions can be drawn:

– Laser cleaning provided the highest lap shear strength for the SikaPower®-880 adhesive and promoted predominantly cohesive failure, indicating the formation of high-quality adhesive joints. Owing to its repeatability, ease of automation, and elimination of hazardous chemical cleaning agents, laser cleaning represents a promising surface preparation method for defense manufacturing.

– Among the conventional mechanical treatments, sanding with P180 abrasive paper consistently provided high lap shear strength for all adhesives investigated and represents a practical alternative when laser equipment is unavailable, particularly for field applications.

– Under identical bonding and surface preparation conditions, X5CrNi18-10 stainless steel exhibited 16–40% higher lap shear strength than EN AW 5754 aluminum alloy. This behavior is attributed to the more favorable surface characteristics of stainless steel, making it a suitable material for load-bearing bonded joints in metal structures.

– Future research should focus on the behavior of adhesively bonded joints under cyclic and dynamic loading, as well as on long-term durability under environmental conditions relevant to defense applications, including high humidity, temperature variations, and UV exposure.

Acknowledgments

This research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-34/2026-03/200109).

References

[1] EBNESAJJAD,S.: *Adhesives Technology Handbook*, 2nd ed., William Andrew Inc., Norwich, NY, USA, 2008.

[2] MARQUES,E.A.S., CARBAS,J.C., AKHAVAN-SAFAR,A., TENREIRO,A.G.F., DA SILVA,F.M.L.: *Structural adhesive bonding in aerospace applications,* Engebook, Edition 1, 2022.

[3] PETRIE,E.M.: *Handbook of Adhesives and Sealants*, McGraw-Hill Companies, New York, USA, 2000.

[4] EBNESAJJAD, S.: *Handbook of Adhesives and Surface Preparation Technology, Applications and Manufacturing,* William Andrew Publishing, 2011.

[5] ROTELLA,G., ALFANO,M., SCHIEFER,T., JANSEN,I.: *Enhancement of static strength and long term durability of steel/epoxy joints through a fiber laser surface pre-treatment*, Int. J. Adhes. Adhes., Vol. 63, pp. 87-95, 2015.

[6] RECHNER,R., JANSEN,I., BEYER,E.: *Influence on the strength and aging resistance of aluminium joints by laser pre-treatment and surface modification*, Int. J. Adhes. Adhes., Vol. 30, pp. 595-601, 2010.

[7] BROCKMANN,W.: *Steel adherends*, in: Durability of Structural Adhesives (Kinloch, A.J., ed.), Elsevier Applied Science, London, pp. Chapter 7, 306, 1983.

[8] EN 1465: *Adhesives-Determination of Tensile Lap-Shear Strength of Bonded Assemblies*, European

Standard, Brussels, Belgium, 2009.

[9] ZDRAVKOVIĆ,N., KLOBČAR,D., MILČIĆ,D., ZUPANČIČ,M., ŽUŽEK,B., MILČIĆ,M., ĐURIĆ,A.: *Influence of Surface Preparation of Aluminum Alloy AW-5754 and Stainless Steel X5CRNI18-10 on the Properties of Bonded Joints*, Materials, Vol. 17(11), 2561, 2024. DOI: 10.3390/ma17112561

[10] ISO 4587: *Adhesive Lap-Shear Strength of Rigid-to-Rigid Bonded Assemblies*, ISO: Geneva, Switzerland, 2003.

[11] HADŽIĆ,A., MOŽE,M., ZUPANČIČ,M., GOLOBIČ,I.: *Superbiphilic Laser-Microengineered Surfaces with A Self-Assembled Monolayer Coating for Exceptional Boiling Performance*, Adv. Funct. Mater. 2023, 34, 2310662.

[12] TA,D., DUNN,A., WASLEY,T., KAY,R., STRINGER, J., SMITH,P., CONNAUGHTON,C., SHEPHARD,J.: *Nanosecond laser textured superhydrophobic metallic surfaces and their chemical sensing applications*,"Appl. Surf. Sci., t. 357, br. Part A, pp. 248-254, 2015.

[13] MARESSA,P., ANODIO,L., BERNASCONI,A., DEMIR,A.G., PREVITALI,B.: *Effect of Surface Texture on the Adhesion Performance of Laser Treated Ti6Al4V Alloy*, J. Adhes., Vol. 91, pp. 518-537, 2015.

[14] YANG,C., MEI,X., TIAN,Y., ZHANG,D., LI,Y., LIU,X.: *Modification of wettability property of titanium by laser texturing*, Int. J. Adv. Manuf. Technol., t. 87, p. 1663-1670, 2016.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

BEHAVIOR OF FLAT SAMPLES MADE OF S355 STRUCTURAL STEEL UNDER STATIC LOADING AND STRESS CONCENTRATION

JOVANA ANTIĆ
Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, Belgrade
email:jrantic@mas.bg.ac.rs

MILETA RISTIVOJEVIĆ
Kontrol Inspekt D.O.O., Belgrade, miletaristivojevic@gmail.com

ALEKSANDAR DIMIĆ
Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, Belgrade,
email: adimic@mas.bg.ac.rs

STEFAN DIKIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Karnegijeva 4. Belgrade,
email: sdikic@tmf.bg.ac.rs

ANA PETROVIĆ
Faculty of Mechanical Engineering, University of Belgrade, Kraljice Marije 16, Belgrade,
email: aspetrovic@mas.bg.ac.rs

SAŠA SPASENIĆ
Kontrol Inspekt D.O.O., Belgrade, spasenic_sale@yahoo.com

SRETEN SPASENIĆ
Kontrol Cert D.O.O., Belgrade, sreten.spasenic@yahoo.com

Abstract: *Machine elements, both simple and complex, are frequently affected by geometric stress concentrations, which significantly influence their load capacity. This is particularly important for components made of structural steels, which are widely used in highly loaded applications. Under overload conditions, unplanned increases in operating load, local stresses at stress concentration areas may reach critical values, including the yield strength and ultimate tensile strength. Available data provides limited information on the influence of geometric stress concentration on the static mechanical properties of materials. In this study, theoretical and experimental investigations are conducted in order to establish a correlation between geometric stress concentration in the form of the notch and static material properties, namely yield strength and ultimate tensile strength. Different notch geometries were analyzed, and corresponding stress–strain (σ–ε) diagrams were obtained through tensile testing of flat samples in laboratory conditions. The results indicate that samples with notches, made of S355 structural steel, have higher values of yield strength and ultimate tensile strength compared to unnotched samples of equivalent cross-sectional area in the notch root. The obtained experimental results can be used for a more precise analysis of the load capacity of machine parts made of S355 with geometric stress concentration.*

Keywords: *geometric stress concentration, strain concentration, yield strength, ultimate tensile strength, S355 structural steel.*

1. INTRODUCTION

Stress concentration is an important factor in the design of machine components, since geometrical discontinuities such as notches, holes, or sharp changes in cross-section cause a non-uniform stress distribution. These local stress increases can strongly influence the mechanical behavior of a component and may lead to premature failure.

Many engineering failures have been linked to stress concentration. A well-known example is the failure of Liberty ships during World War II, where structural fractures originated at welded joints and geometrical discontinuities [1]. However, similar effects occur much more frequently in engineering practice, where they may not lead to catastrophic failure but can still significantly affect the performance and reliability of components. Consequently, a considerable number of studies and experimental investigations have been devoted to the analysis of stress concentration and notch effects on mechanical properties. Various sample geometries have been employed, including cylindrical and flat samples, as well as specially designed geometries aimed at controlling the stress state, such as butterfly-type samples [2]. In addition, different notch configurations have been

Corresponding author: J. Antić;
E-mail: jrantic@mas.bg.ac.rs
DOI: 10.5937/OTEH26113A

investigated, such as V-shaped, U-shaped, semicircular and key-hole notches [3,4]. Furthermore, various materials have been considered in order to evaluate their influence on stress concentration and notch effects.

In the present study, the influence of geometry on the stress state and mechanical behavior of S355 structural steel was analyzed. Structural steel S355 is widely used in engineering applications, particularly for load-bearing and safety-critical components. Due to its frequent use in structures exposed to significant loads, it is essential to understand how stress concentration influences its mechanical properties. Special attention is given to notch yield strength (σ_T) and notch ultimate tensile strength (σ_m), as these represent critical points in the material response.

Four different flat sample geometries were considered, including one smooth sample and three samples with different notch geometries. The aim of this work is to investigate the effect of notch geometry on yield strength and ultimate tensile strength and to better understand the role of stress concentration in structural steel components.

2. MATERIALS AND METHODS

2.1. Material

The samples used in this study were made of structural steel S355. According to [5], the minimum value of yield strength (R_e) for this material is 355 MPa, while ultimate tensile strength (R_m) is specified between 470 and 630 MPa. As defined in [7], R_e represents the maximum value of stress prior to the first decrease in force in uniaxial conditions, while R_m is the stress corresponding to the maximum force.

The samples were initially cut using a water jet process into plates of dimensions 22 × 210 × 8 mm. The final sample geometry, including the notch profile, was obtained by CNC milling. The end surfaces were subsequently ground to achieve a uniform thickness of 6 mm with high accuracy.

2.2. Sample Geometry

Tensile tests were performed on flat samples. The general sample geometry is shown in Figure 1, where the notch radius (R) and the distance between notch centers (D) are defined as the varying parameters.

Four different geometries were considered by varying R. The minimum cross-sectional area at the notch root was kept constant and equal to that of the smooth sample. As a result, D varied accordingly.

For each geometry, the stress concentration factor (α_k) was determined analytically using expressions provided in literature [6]. The factor α_k is defined as the ratio between the maximum stress at the notch root (σ_{max}) and the nominal stress (σ_{nom}):

$$\alpha_k = \frac{\sigma_{max}}{\sigma_{nom}} \tag{1}$$

where σ_{nom} is calculated from the applied force and the minimum cross-sectional area at the notch root. Due to the presence of the notch, the stress at the notch root is locally increased (Figure 2), and the corresponding σ_{max} can be determined using Equation 1.

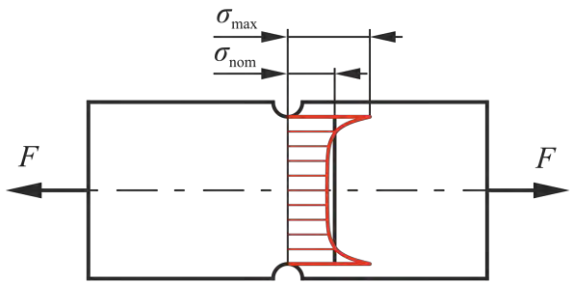


Figure 2 Stress distribution in the notch region under tensile loading

The values of R, D and α_k are summarized in Table 1. The first sample was smooth (without a notch), whereas the remaining samples contained notches with progressively decreasing radii.

Table 1. Values of R, D, and α_k for different sample geometries

	R [mm]	D [mm]	α_k [-]
E1	0 (smooth sample)	-	1
E2	3.75	20	1.69
E3	1.25	15	2.38
E4	0.625	13.75	3.07

2.3. Experimental Procedure

Tensile tests were performed using an Shimadzu testing machine with a 250 kN capacity at the Faculty of Technology and Metallurgy, University of Belgrade (Figure 4). Testing was carried out under displacement control, with a constant crosshead separation rate of 5 mm/min. The experimental procedure was conducted in accordance with [7].

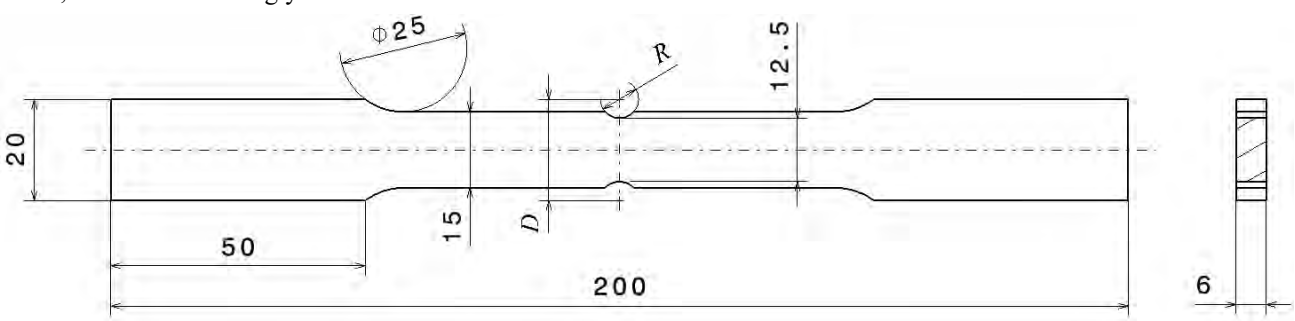


Figure 1 Geometry and dimensions of the tested specimen

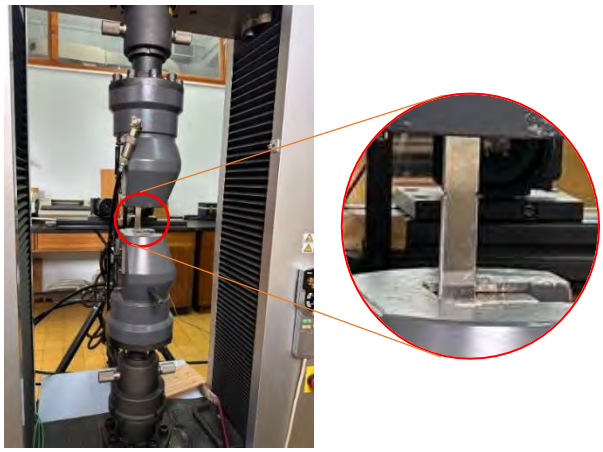


Figure 4. Tensile testing machine with mounted sample

All tests were performed at room temperature. Force and crosshead displacement were recorded during testing. Strain was determined from the crosshead displacement, while the engineering stress $\sigma(t)$ was calculated as:

$$\sigma(t) = \frac{F(t)}{A_0} \tag{2}$$

where $F(t)$ is the applied force at time and A_0 is the initial area at the narrowest section of the sample (12.5 × 6 mm, as shown in Figure 1). The calculated stress values were then paired with the corresponding strain to construct the engineering stress–strain curves.

3. RESULTS

The fractured samples for all tested geometries are shown in Figure 5. The samples were aligned from one side in order to enable a direct comparison of their elongation after fracture. A noticeable difference in elongation can be observed between the smooth and notched samples. The smooth sample exhibits the highest elongation, while only minor differences in elongation are observed among the notched samples.

Figure 5. Fractured samples after tensile testing

The fracture surfaces of the tested samples are shown in Figure 6. The morphology indicates ductile fracture, characterized by pronounced material tearing. However, the extent of plastic deformation is reduced in the notched samples compared to the smooth sample.

Figure 6. Fracture surfaces of tested samples

Engineering stress–strain (σ–ε) curves obtained from tensile tests for all four sample geometries are presented in Figure 7.

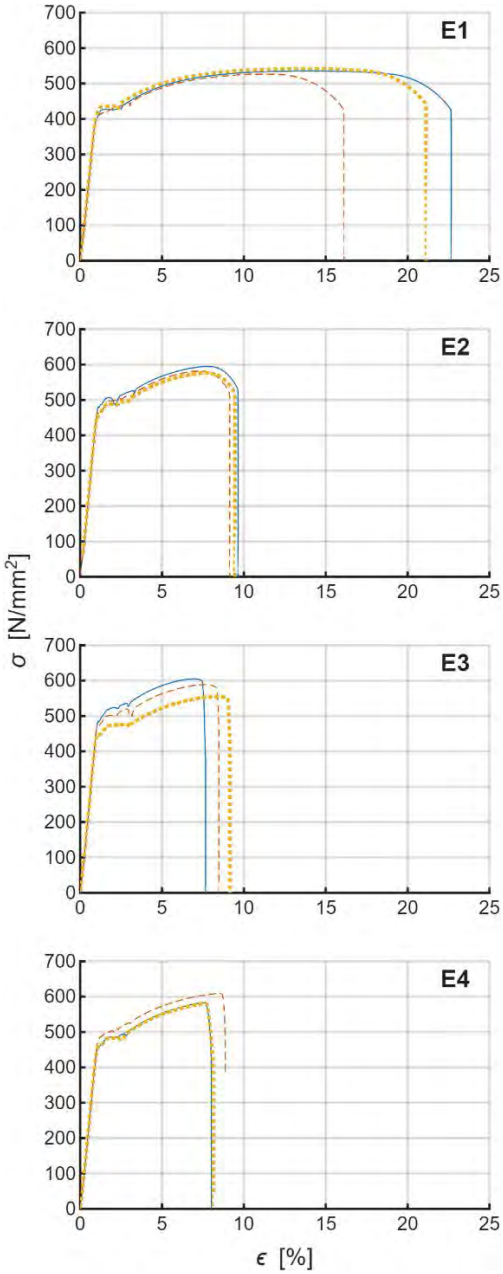


Figure 7. Stress–strain diagrams for four sample geometries (E1–E4), with three samples per geometry

The curves exhibit similar overall behavior, with a clear elastic region followed by yielding and plastic deformation up to fracture. For most geometries, a good agreement between the curves can be observed. Although one of the smooth samples exhibited noticeably lower elongation

compared to the others, the overall repeatability of the results was satisfactory. The value reported in the paper was obtained from the testing-machine data and is based on the crosshead displacement, rather than on the elongation measured directly on the specimen. It was used only for a relative comparison among the tested specimens. It does not represent the percentage elongation after fracture A, as defined by the tensile-testing standard, and a direct comparison of this value with the standard requirement for A would therefore not be appropriate.

4. DISCUSSION

The average values obtained from the three smooth samples are:

$$R_e = 422.8\frac{\mathrm{N}}{\mathrm{mm}^2};\ R_m = 534.5\frac{\mathrm{N}}{\mathrm{mm}^2};\ \varepsilon_{\max,1} = 19.98\%. \quad (3)$$

Table 2 summarizes the average values obtained from three samples for each notched geometry.

Table 2. Average values of σ_T, σ_m, and ε_{max} for notched samples and their corresponding ratios to the smooth-sample values

	σ_T	$\frac{\sigma_T}{R_e}$	σ_m	$\frac{\sigma_m}{R_m}$	ε_{max}	$\frac{\varepsilon_{max}}{\varepsilon_{max,1}}$
E2	467.0	1.105	584.5	1.094	9.414	0.47
E3	467.7	1.106	582.8	1.090	8.434	0.422
E4	468.7	1.109	591.0	1.106	8.341	0.417

The measured value of R_e exceeds the minimum value specified by the standard by approximately 19%, while the corresponding value of R_m falls within the prescribed range.

The obtained results show that, for all notched samples, both σ_T and σ_m exceed the corresponding values of R_e and R_m, respectively. The values of σ_T are from 10.5% to 10.9% higher than R_e, while σ_m exceeds by 9.4% to 10.6%. The differences between individual notch geometries are very small, remaining below 1% for σ_T, whereas the variation in σ_m slightly exceeds 1%. This indicates a limited influence of notch geometry on strength within the investigated range for tested material.

In addition to the increase in strength, a significant reduction in elongation is observed for notched samples compared to the smooth sample. The elongation of the notched samples is reduced to 42–47% of the smooth sample elongation.

The literature shows that the effect of stress concentration on mechanical properties is not unambiguous. In some cases, the triaxial stress state in the notch region results in lower values of σ_T and σ_m compered to the R_e and R_m due to early crack initiation at stress concentration sites, corresponding to notch weakening [8]. However, cases of increased strength have also been reported. In such cases, the material in the notch region is constrained in its deformation, as its elongation is restricted by the surrounding material. This constraint leads to the development of tangential and radial stresses, resulting in a triaxial stress state, which requires higher axial stress levels for yielding and plastic deformation [9]. As a result, an apparent increase in strength, known as notch strengthening, is observed, which is consistent with the results obtained in the present study.

Similar behavior has been reported in [10], where tensile tests were performed on cylindrical samples made of 316L(N) stainless steel with symmetric double U-notches. Five different U-shaped notches with varying radii were considered, and the results showed a clear increase in both σ_T and σ_m as the notch radius decreased.

Similar findings were reported in [11], where tensile tests on cylindrical cast iron samples with V-shaped notches showed that the maximum strength of the notched samples exceeded that of smooth samples. This behaviour was attributed to the increased stress triaxiality in the notch region, with the notch strengthening effect becoming more pronounced as the level of stress triaxiality increased.

Most experimental studies reported in the literature are conducted on cylindrical samples, whereas the present study focuses on flat samples. Although the phenomenon of notch strengthening has been widely investigated, its quantitative characterization for flat samples remains limited. In this context, the present results contribute to a better understanding and quantification of notch strengthening in flat sample geometries.

For the investigated sample geometry and considered levels of stress concentration, the increase in σ_T and σ_m for all notched samples (E2–E4) relative to the R_e and R_m was clearly quantified. Conventional expressions for elastoplastic materials indicate that notch yield strength and notch ultimate tensile strength are close to those of the smooth sample [12]:

$$R_e \approx \sigma_T,\quad R_m \approx \sigma_m. \quad (4)$$

The results of this study confirm that these expressions are on the side of safety and can be applied as such in engineering practice. For more accurate calculations, the results presented in this study can be used.

5. CONCLUSION

The results of this study show that all notched samples exhibit higher values of σ_T and σ_m compared to the R_e and R_m, which can be attributed to the phenomenon known as notch strengthening. The increase in strength is approximately 10%, while the differences between individual notch geometries remain below 1% for the tested geometries.

A significant reduction in elongation is observed in the notched samples, with values reaching only 42–47% of those measured for the smooth sample.

The results further indicate that, within the investigated range, the variation in notch geometry has a limited influence on strength, while its effect on ductility is more pronounced for the tested material.

ACKNOWLEGEMENTS

This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under Contract No. 451-03-34/2026-03/200105 dated February 5, 2026 and Contract No. 451-03-34/2026-03/ 200135. The authors also acknowledge the support of the companies Kontrol Inspekt and EDePro, which contributed to the realization of this research.

References

[1] ZHANG, W.: *Technical Problem Identification for the Failures of the Liberty Ships,* Challenges, 7 (2) (2016)

[2] DUNAND, M., MOHR, D.: *Optimized butterfly specimen for the fracture testing of sheet materials under combined normal and shear loading*, Engineering Fracture Mechanics, 78 (2011) 2919–2934.

[3] SOLOUKI, A., ALIHA, M.R.M., MAKUI, A., CHOUPANI, N.: *Analyzing the effect of notch geometry on the impact strength of 3D-printed specimens*, Materials Testing, 65 (2023).

[4] SHA, Z., TENG, Y., POH, L.H., PEI, Q., XING, G., GAO, H.: *Notch strengthening in nanoscale metallic glasses*, Acta Materialia, 169 (2019) 147–154.

[5] EN 10025-2: Hot rolled products of structural steels – Part 2: Technical delivery conditions for non-alloy structural steels.

[6] PILKEY, W.D., PILKEY, D.F., BI, Z.: Peterson's Stress Concentration Factors, 4th ed., John Wiley & Sons, Hoboken (2020).

[7] SRPS EN ISO 6892-1:2020: Metallic materials – Tensile testing – Part 1: Method of test at room temperature.

[8] NIU, H., SUN, Y., LIN, C.: *Study on notch tensile properties of Fe-Mn-Si shape memory alloy*, Materials Research Express, 11 (2024).

[9] DIETER, G.E.: *Mechanical Metallurgy*, McGraw-Hill, New York, 1961.

[10] GANESH KUMAR, J., NANDAGOPAL, M., PADMANABHAN, P., LAHA, K., MATHEW, M.D.: *Effect of notch root radius on tensile behaviour of 316L(N) stainless steel*, Materials at High Temperatures, 31 (2014) 239–248.

[11] IKEDA, T., NODA, N.-A., SANO, Y.: *Conditions for notch strength to be higher than static tensile strength in high-strength ductile cast iron*, Engineering Fracture Mechanics, 206 (2019) 75–88.

[12] MITROVIĆ, R., RISTIVOJEVIĆ, M., ROSIĆ, B.: *Mašinski elementi 1*, 2nd ed., Univerzitet u Beogradu, Mašinski fakultet, Belgrade (2021).

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

TEMPORAL VARIATIONS IN BACKGROUND RADIATION AND THEIR CORRELATION WITH METEOROLOGICAL PARAMETERS IN VRANJE, SERBIA

MILAN TANIĆ
The Academy of Applied Preschool Teaching and Health Studies, Department of Technical and Technological Studies Kruševac, Kruševac, milantanic@yahoo.com

DENIS DINIĆ
University of Belgrade, Faculty of Technology and Metallurgy, Belgrade, denis.dinic@yahoo.com

MLADEN NIKOLIĆ
The Academy of Applied Preschool Teaching and Health Studies, Department of Technical and Technological Studies Kruševac, Kruševac, mladennikolic2603@gmail.com

STEVAN STUPAR
University of Criminal Investigation and Police Studies, Belgrade, stevan.stupar13@gmail.com

Abstract: *The temporal variability of the ambient dose equivalent rate, $\dot{H}^*(10)$, in Vranje, Serbia, and its relationship with meteorological parameters during 2024 were investigated. Hourly $\dot{H}^*(10)$ data reported by an automatic radiation monitoring post within the early warning network and calculated daily $\dot{H}^*(10)$ were analyzed against daily values of selected meteorological variables using descriptive statistics, nonparametric tests, simple correlation and multiple linear regression analyses. A statistically significant seasonal effect was confirmed, while diel differences were consistent but environmentally small. Within the hourly $\dot{H}^*(10)$ dataset, multi-hour spikes were identified and were associated with rainfall events. Correlation analysis indicated that temperature-related variables and atmospheric moisture were positively associated with $\dot{H}^*(10)$, while pressure, cloud cover, and precipitation showed inverse relationships. The multiple regression model explained a moderate proportion (27.9 %) of the daily variability of $\dot{H}^*(10)$, suggesting that meteorological factors contribute substantially, but not entirely, to the observed fluctuations. Improved understanding of background radiation dynamics at local scale under varying weather conditions enables robust interpretation of environmental monitoring data and strengthens the reliability of the early warning network.*

Keywords: *dose rate, seasonal variation, monitoring, radionuclide, weather.*

1. INTRODUCTION

Natural background radiation is an omnipresent part of the environment. The ambient gamma radiation mainly originates from terrestrial radionuclides (decay series of ^{238}U and ^{232}Th including short-lived radon progeny, and ^{40}K), cosmic radiation and cosmogenic radionuclides. Anthropogenic contributions to the background radiation are much smaller and mainly reflect fallout from atmospheric nuclear weapons testing and contamination from nuclear accidents [1].

Because a meaningful fraction of the near-surface gamma signal is linked to radon progeny and atmospheric mixing, the background level is not constant. Rain can cause short-term increases in dose rate by washing out aerosol-bound radionuclides, particularly radon progeny, from the atmosphere and depositing them on the ground, followed by decay toward baseline [2]. Temperature, humidity, pressure, wind, cloudiness, and boundary-layer stability modulate radon exhalation, dispersion, and deposition. Conversely, snow cover attenuates terrestrial radiation [3]. Recent studies report diurnal and seasonal structures in dose-rate time series and statistically detectable associations with meteorological variables [2, 6].

This variability is operationally important for early warning networks. Automatic stations continuously monitor the ambient dose equivalent rate, $\dot{H}^*(10)$ (an operational quantity used for area monitoring) to provide rapid indication of a nuclear or radiological emergency. In Serbia, the national early warning system is operated by the Serbian Radiation and Nuclear Safety and Security Directorate (SRBATOM). The legacy network was established in 2007 with nine stations, while the 2023 modernization expanded coverage to 33 automatic radiation monitoring posts (ARMP), including Vranje, which was not part of the earlier system.

Vranje was selected due to local and societal motivations. Southern Serbia was among the areas investigated for potential environmental consequences of depleted uranium (DU) used during the 1999 NATO campaign, and DU-related monitoring and remediation have been reported for sites in the wider region [7]. Vranje is often

Corresponding author: M. Tanić;
E-mail: milantanic@yahoo.com
DOI: 10.5937/OTEH26114T

mentioned in public discourse in connection with perceived increases in cancer burden, which makes transparent, data-driven radiation monitoring relevant for risk communication and public trust [8].

The aims of this study were: (i) to quantify variability of the ambient dose equivalent rate in Vranje, with emphasis on seasonal and diurnal contrasts; and (ii) to evaluate the influence of meteorological parameters on dose-rate variability, including multivariable regression.

This work is relevant for nuclear security: improved characterization of "normal" variability can reduce unnecessary mobilization and analyst fatigue, while preserving sensitivity and credibility when a genuine radiological threat or release occurs, especially when small anomalies may be partially masked by weather-driven background shifts. These results can support interpretation within the national early warning system by helping distinguish meteorologically induced fluctuations from event-like anomalies that might trigger alarms [9].

2. MATERIALS AND METHODS

Study area

Vranje is a town in southern Serbia (42°33' N, 21°55' E, 430 m a.s.l), near the borders with North Macedonia and Bulgaria. The area of Vranje Basin is characterized by an assembly of crystalline schists, gneisses, and granitoids of Paleozoic age, overlain locally by Neogene-Quaternary sediments. Alluvial deposits are developed along the South Morava and its tributaries. Surrounding elevations are composed predominantly of metamorphic and igneous formations. The dominant soil types are alluvial soils (Fluvisols) and colluvial deposits in valley bottoms. Cambisols and Leptosols are prevalent on surrounding slopes and hilly terrain. According 2022 Census, the City of Vranje has approximately 74000 inhabitants.

Vranje has a moderate continental climate, with hot summers and fairly cold winters. Based on climate data from Republic Hydrometeorological Service of Serbia (RHSS) (1981 to 2010) a mean annual air temperature is 11.1 °C, with cold winters (average of –0.1 °C in January) and warm summers (21.6 °C in July and August). Total annual precipitation is 578.3 mm, with wetter spring and drier late summer. Prevailing winds are northeasterly, with low to moderate speeds (2.2–4.2 m s^{-1}). Total sunshine hours are 2123.9 h.

Based on the available data, no nuclear or radiological emergencies were reported in 2024 within the area surrounding the ARMP. The closest nuclear facilities are the Vinča Nuclear Institute in Belgrade (app. 270 km, straight-line) and the Kozloduy Nuclear Power Plant in Bulgaria (app. 200 km). Therefore, it can be reasonably concluded that the Vranje ARMP was not influenced by potential nuclear effluents during the reporting period.

Data Collection

Datasets were collected from publicly available sources. Hourly averaged values of $\dot{H}^*(10)$ were retrieved from the European Radiological Data Exchange Platform (EURDEP) from January 1 to December 31, 2024 (station RS31149) [10]. Meteorological parameters identified as a key relevant drivers influencing ambient background radiation [11], namely daily mean of atmospheric pressure (p), air temperature (T), minimum air temperature at 5 cm above the ground (T_{5cm}), relative humidity (RH), vapor pressure (e), sunshine hours (S), cloud cover (N), wind speed (U), and precipitation (R), were taken from the Meteorological Annual for 2024 [12].

The Main Meteorological Station (MMS) Vranje is located at 42º33' N latitude and 21º55' E longitude, at 433 m a.s.l. The ARMP is permanently installed within the boundaries of the MMS Vranje, and is equipped with the MIRA dose monitoring system from Scienta Envinet. MIRA measures $\dot{H}^*(10)$ with two temperature and background compensated Geiger–Muller tubes: in the low-dose (LD) range from 10 nSv h^{-1} to 1 mSv h^{-1} with a sensitivity of 823 count · h · min^{-1} · μSv, and in high-dose (HD) range from 0.01 mSv h^{-1} to 10 Sv h^{-1}.

Measurement uncertainty of the $\dot{H}^*(10)$ in the LD range was evaluated following Guide to the expression of uncertainty in measurement principles (JCGM GUM-1:2023). The manufacturer-specified accuracy of ± 15 % was taken as a Type B uncertainty component with rectangular distribution, which corresponds to a relative standard uncertainty of 8.66 %. The statistical component was estimated from counting statistics assuming Poisson behavior of the detector for hourly averaged data. Within the observed dose rate range in this work, the relative standard uncertainty (Type A) due to counting statistics fell in the range 0.8–1.2 %. The combined relative standard uncertainty was ~8.7 %. The expanded relative uncertainty at 95 % confidence level ($k = 2$) of the reported values was about 17.5 %.

Statistical analysis

The Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests assessed normality of the $\dot{H}^*(10)$ and meteorological data at a significance level of 0.05. Outliers in $\dot{H}^*(10)$ was performed using mean absolute deviation (MAD) approach, and all observations with absolute value of z-score higher than 3.5 were treated as outliers.

Statistical analyses were conducted using IBM SPSS 29.0. Correlation and linear regression analyses assessed relationships among $\dot{H}^*(10)$ and the meteorological records. OriginPro 8.5 was utilized for plotting.

The Kruskal-Wallis (K-W) test, followed by the Mann-Whitney (M-W) U test, was applied to the original variables to assess whether the season (winter, spring, summer, autumn) or diel period (day, night) influenced the variability of the $\dot{H}^*(10)$. For the K-W test, a significant difference was recognized at $p < 0.05$, while for the M-W test, p value was set after the Bonferroni correction based on the number of pairwise comparisons (six tests, adjusted threshold $p < 0.001$).

Seasons for the georeferences of the ARMP Vranje were

determined using the 2024 astronomical equinox and solstice dates. Sunrise and sunset times were calculated based on zenith angle 90.833°, accounting for atmospheric refraction and daylight-saving time adjustments.

3. RESULTS AND DISSCUSION

Ambient dose equivalent rate

The number (n) of 8784 hourly averaged $\dot{H}^*(10)$/nSv h^{-1} values was reported by the MIRA station RS31149 in 2024. The descriptive statistics of $\dot{H}^*(10)$ for the total number of observations are presented in the Table 1. The $\dot{H}^*(10)$ values ranged from 82.74 to 260.50 nSv h^{-1} with a mean of 101.32 nSv h^{-1}. The coefficient of variation (CV) which is in a good indicator in environmental studies of the presence of external impacts, was rather small (7.75 %), which highlights the absence of any major external influence on the natural background radiation.

The K-S test of normality (0.086, $p < 0.001$), along with the evaluation of the Q-Q plot, confirmed a pronounced non-normal distribution of $\dot{H}^*(10)$, with a heavy upper tail. A skewness of 3.65 and kurtosis of 37.73 indicated a highly right-skewed and leptokurtic distribution.

The MAD method for outlier detection, which is a preferred robust method for heavy-tailed, non-normally distributed datasets, identified 109 outliers in total (6 in winter, 16 in spring, 51 in summer, and 36 in autumn). A closer inspection of the detected outliers showed that these values were not isolated but formed multi-hour clusters, and it was confirmed that all of them, without exception, were recorded on rainy days.

Table 1. Descriptive statistics of the ambient equivalent dose rate, $\dot{H}^*(10)$/nSv h^{-1} by season and diel cycle

Season	Period	n	Mean	SD	CV/%	Median	Min	Max	Kurtosis	Skewness
Winter	Day	912	96.60	4.22	4.37	96.19	82.74	129.01	9.03	1.80
	Night	1241	97.13	4.66	4.80	96.68	85.99	126.37	7.53	1.89
	Season	2153	96.90	4.49	4.63	96.44	82.74	129.01	8.12	1.87
Spring	Day	1312	99.15	5.39	5.43	98.32	88.80	158.50	19.19	2.74
	Night	915	100.85	5.58	5.53	100.12	89.64	145.04	12.40	2.29
	Season	2227	99.85	5.53	5.54	99.11	88.80	158.50	15.37	2.47
Summer	Day	1323	105.62	8.55	8.10	105.19	89.86	260.50	96.28	6.71
	Night	925	108.69	8.62	7.93	108.57	91.86	194.03	23.56	3.21
	Season	2248	106.89	8.71	8.15	106.44	89.86	260.50	60.69	5.02
Autumn	Day	910	100.93	8.92	8.84	99.32	89.93	183.47	28.18	4.13
	Night	1246	101.86	7.53	7.39	101.07	89.08	185.95	26.46	3.63
	Season	2156	101.47	8.15	8.04	100.34	89.08	185.95	27.82	3.89
Annual	Day	4457	100.91	7.84	7.77	99.65	82.74	260.50	57.77	4.62
	Night	4327	101.75	7.85	7.71	100.21	85.99	194.03	18.09	2.68
	Year	8784	101.32	7.85	7.75	99.95	82.74	260.50	37.73	3.65

Meteorological data

The mean air pressure recorded at MMS Vranje in 2024 was 966.5 mb, the relative humidity 73 %, the air temperature 13.6 °C, the minimum air temperature at 5 cm above the ground 5.6 °C, the vapor pressure 11.5 mb, the wind speed 1.5 m s^{-1} and the cloud cover 4.7 tenths. There were a total of 2393 h of sunshine hours, 121 calm days, 715 mm of precipitation, 115 rainy days, 12 snowy days and 6 days with snow cover.

Temporal variation

As shown in Table 1, the highest values of $\dot{H}^*(10)$ were recorded in summer, followed by autumn, with mean values of 106.89 and 101.47 nSv h^{-1}, respectively. The lowest $\dot{H}^*(10)$ was measured in winter (mean 96.90 nSv h^{-1}), followed by a slight increase in spring, when the average value was 99.85 nSv h^{-1}. The same seasonal order was reported by other similarly designed studies in South Korea, Poland and Romania [2, 5, 13]. Based on the SD and CV values, the variability of the dose rate was in the same order: summer > autumn > spring > winter. Dose rates were slightly, but consistently higher during the night than during daylight.

Diurnal variation in gamma radiation is commonly linked to diurnal fluctuations in atmospheric radon concentrations and the formation of nighttime temperature inversions, which lead to a decrease in gamma radiation after sunrise and a progressive rise after sunset [2]. This interpretation is supported by a study reporting that the diurnal pattern was observed only in gamma energies originating from terrestrial background radioactivity, majority from the ^{214}Bi [6]. A similar pattern was observed in our study. Figure 1 presents mean values of $\dot{H}^*(10)$ in each hour for different seasons. It is evident that across all seasons, $\dot{H}^*(10)$ showed a clear diurnal pattern. In that regard, winter was the most stable, with mean hourly $\dot{H}^*(10)$ varying slightly from 96 to 99 nSv h^{-1}, with a weak decline at midday. $\dot{H}^*(10)$ increased in spring from 98 to 102 nSv h^{-1} and showed a daytime minimum followed by evening recovery. Summer had the highest $\dot{H}^*(10)$ values, from 104 to 111 nSv h^{-1}, and the strongest cycle, peaking in early morning (02:00 to 06:00 h), dropping toward midday when the lowest values were reported (11:00 to 15:00 h), then rising again at evening and night. Autumn is intermediate, with morning

maximum (07:00 h, 105 nSv h^{-1}) and midday minimum (15:00, 98 nSv h^{-1}). The calculated *SD* indicated larger variability in summer and autumn than in winter and spring.

Monthly averages showed in Figure 2 confirmed the existence of seasonal variability. Values were lowest from January to March. An increase occurred during spring and early summer, with a peak in August. The period September–October was characterized by a steady decline. A smaller secondary rise was observed in November, followed by a decline in December. Amestoy at al. [14] attributed such seasonal variability to changes in surface soil moisture affecting gamma rays primarily originating from Th and K. They reported that increases in soil moisture of 25 % at the end of summer and 43 % at the beginning of spring corresponded to decreases in gamma radiation from these two radionuclides by the same percentages.

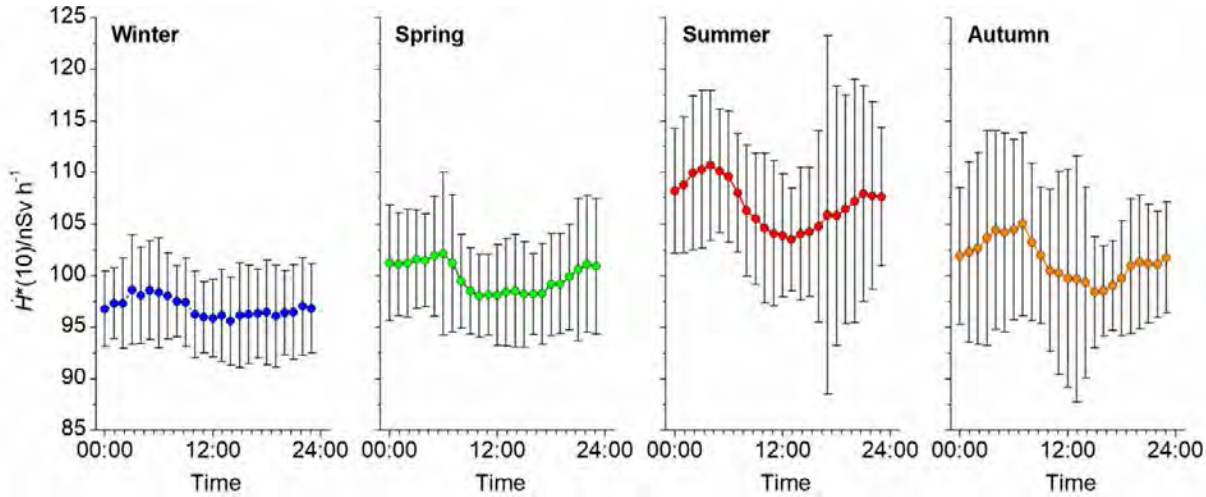


Figure 1. Diurnal variation of the ambient equivalent dose rate by seasons (error bars indicate ±*SD*)

Based on Table 1 and Figure 2, nighttime $\dot{H}^*(10)$ mean values were consistently higher than during daylight (by approximately 1 %). The difference was most pronounced during the summer months and was the highest in June (3.5 %). The exception was noted in November, when the mean dose rate during the day exceeded the nighttime value by 2 %. Such nocturnal enhancement of the ambient dose occurs as radon builds up in the stable boundary layer [15].

The K-W test confirmed a statistically significant seasonal effect ($\chi^2 = 2644.707$, $df = 3$, $p < 0.001$) on $\dot{H}^*(10)$. The mean ranks increased from winter (2606.29) to spring (3960.93) and autumn (4454.47), while the highest values were obtained in summer (6471.33).

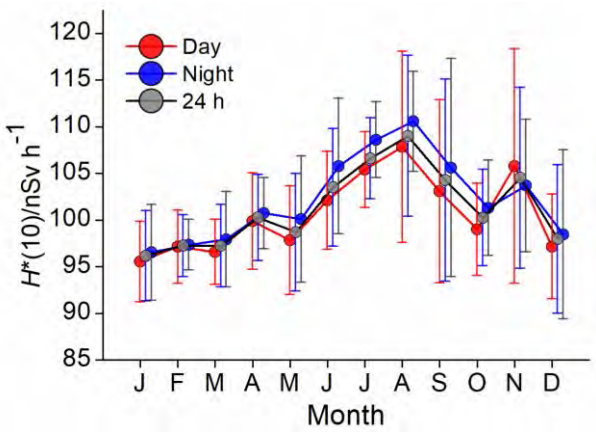


Figure 2. Monthly variation of the ambient equivalent dose rate by diel period (error bars indicate ±*SD*)

Pairwise M-W tests (Table 2) confirmed statistically significant differences for all seasonal pairs ($p < 0.001$). However, the strength of the contrasts was not uniform. The largest separation quantified by the effect size (*ES*) calculated as the quotient of the *Z* value and the square root of the number of observations in the comparison, was observed between winter and summer ($ES = 0.69$), and between spring and summer ($ES = 0.53$). Comparisons involving autumn were mainly moderate (with summer $ES = 0.42$; with winter $ES = 0.38$), while difference between spring and autumn was small ($ES = 0.11$).

For the diel cycle (Table 2), a day–night difference was detected ($Z = -5.676$, $p < 0.001$), but the *ES* was negligible. Therefore, although diel modulation is statistically detectable, it is likely not environmentally important compared with the seasonal effect.

Correlation and linear regression analysis

Since meteorological data at 1 h-resolution were not available and only daily averages of selected meteorological parameters were accessible, mean daily $\dot{H}^*(10)$ values were calculated as the averages of 24 hourly values (from 00:00 to 23:00 h) and further used for the purposes of correlation and linear regression analysis.

Since the S-W test showed that all variables included in the analysis did not conform to a normal distribution, Spearman's correlation coefficient was chosen to quantify interrelations between them. The results are presented in Table 3. All correlations were statistically significant (confidence level 99 %) except for wind speed. According to the classification proposed by Joo et al. [5] the daily

mean air temperature, air temperature at 5 cm above the ground, vapor pressure and sunshine hours were moderately and positively related to the mean daily $\dot{H}^*(10)$, whereas atmospheric pressure, relative humidity, cloud coverage, and precipitation showed weak inverse associations.

Table 2. Mann-Whitney U test statistics for pairwise seasonal and diel comparisons of $\dot{H}^*(10)$

Category	*n*	Ranks		Mann-Whitney U	Wilcoxon W	*Z*	*p*	*ES*
		Mean Rank	Sum of Ranks					
Day	4457	4241.18	18902926.00	8968273.00	18902926.00	-5.676	0.000	0.06
Night	4327	4548.37	19680794.00					
Winter	2153	1782.89	3838556.00	1519775.00	3838556.00	-20.976	0.000	0.32
Spring	2227	2584.57	5755834.00					
Winter	2153	1298.98	2796704.50	477923.50	2796704.50	-46.090	0.000	0.69
Summer	2248	3064.90	6889896.50					
Winter	2153	1678.43	3613649.50	1294868.50	3613649.50	-25.129	0.000	0.38
Autumn	2156	2630.91	5672245.50					
Spring	2227	1544.87	3440423.50	959545.50	3440423.50	-35.721	0.000	0.53
Summer	2248	2924.66	6574626.50					
Spring	2227	2059.49	4586483.00	2105605.00	4586483.00	-7.046	0.000	0.11
Autumn	2156	2328.87	5021053.00					
Summer	2248	2730.77	6138774.00	1235790.00	3561036.00	-28.155	0.000	0.42
Autumn	2156	1651.69	3561036.00					

The results of the correlation analysis show that temperature-related variables were the main meteorological factors influencing the environmental dose rate. This can be explained by the fact that higher temperatures reduce soil moisture and empty soil pores, which in turn increase the dose rate in the air [2]. On the other hand, temperatures below 0 °C lead to soil freezing, which at the very least hinders gas exchange with the atmosphere and consequently limits the migration of radon into the air [3]. The positive correlation with sunshine hours together and the negative correlation with cloud cover result from the coincidence of sunny weather with high-pressure, low-wind, and stable atmospheric mixing conditions, which are characteristic of a moderate continental climate and promote near-surface radon accumulation [11].

Table 3. Spearman's correlation coefficient between mean daily $\dot{H}^*(10)$ and meteorological parameters (significant at the 0.01 level; *nonsignificant correlation)

Para.	Coeff.	Para.	Coeff.	Para.	Coeff.
p	–0.26	*RH*	–0.34	*N*	–0.21
T	0.57	*e*	0.53	*U*	–0.09*
T_{5cm}	0.51	*S*	0.33	*R*	–0.24

An inverse association with atmospheric pressure is probably a consequence of inhibited emission of radioactive gases from the soil under conditions of high atmospheric pressure [2]. In addition, photon intensity is inversely proportional to atmospheric pressure, which also influences the gamma dose rate [9]. Also, the reason for the negative association with atmospheric pressure may lie in the barometric modulation of secondary cosmic radiation, as greater atmospheric mass increases shielding and reduces the cosmic-ray contribution at ground level as well as reduction of cosmogenic radionuclides generation (e.g. ^{7}Be, ^{14}C) [11, 14]. Since it is well established that wet deposition of radon progeny present in the atmosphere, particularly ^{214}Pb and ^{214}Bi, leads to an increase in ambient gamma dose rate [2], a positive correlation between gamma dose rate and precipitation would be expected. The negative value of Spearman's correlation coefficient for precipitation initially appears to contradict the observed hourly spikes during rain. However, this is explained by the dual nature of precipitation's impact: while "washout" causes immediate, short-lived increases in $\dot{H}^*(10)$, the resulting increase in soil moisture over a 24-hour period reduces radon emanation and attenuates terrestrial gamma radiation, leading to a net decrease in the daily mean [2]. Although the correlation with wind speed was not statistically significant, the negative coefficient suggests that under conditions of weak wind flow, atmospheric mixing is reduced and air masses may become relatively stagnant. This limited dispersion can promote the accumulation of radionuclides near their release or resuspension sources, thereby contributing to elevated ambient dose rates in the surrounding area [11].

Multiple linear regression analysis quantified the meteorological influence on the $\dot{H}^*(10)$ daily mean using the dataset of daily averaged meteorological parameters ($n = 366$). Meteorological parameters are interconnected as a change in one variable can affect others [11]. To avoid multicollinearity among thermodynamic variables (p, T, T_{5cm}, e), vapour pressure was selected as a representative measure of atmospheric moisture, while sunshine duration was used as a radiative descriptor rather than N. The explanatory model given by regression Eq. (1) explained 27.9 % of the variance in $\dot{H}^*(10)$ daily averages which is below the explanatory power reported in some studies [11]. All predictors kept in the analysis contributed significantly to the model ($p < 0.05$), and based on the variance inflation factor ($VIF < 3.5$), multicollinearity was acceptable. Although precipitation showed a statistically significant monotonic association with $\dot{H}^*(10)$ based on the Spearman's correlation coefficient, it did not exhibit a significant linear effect in the multiple regression and did not improve model

performance. This confirms that precipitation effects are likely nonlinear and episodic at the daily scale.

$$\dot{H}^*(10) = 107.233 + 0.610e - 0.129RH - 0.301S - 0.911U \quad (1)$$

Given the standardized regression coefficients, β, vapour pressure (e) showed the strongest positive association (0.479), indicating that higher absolute humidity is associated with increased $\dot{H}^*(10)$ which is directly linked to the air temperature, and precipitation. Relative humidity (RH) displayed a moderate negative effect (–0.331) when vapour pressure was included in the model, indicating a conditional moisture-related influence. Sunshine hours, S, (–0.227) and wind speed, U, (–0.126) were negatively associated with $\dot{H}^*(10)$, which is consistent with enhanced atmospheric mixing and dilution under increased solar radiation and wind [2]. The Durbin–Watson statistic (0.89) indicated positive serial correlation, reflecting the time-series nature of daily observations. Therefore, results of the regression should be interpreted as associations rather than independent effects.

4. CONCLUSION

This study, based on a one-year time series, demonstrates that variations in the $\dot{H}^*(10)$ in Vranje exhibit distinct seasonal and diurnal patterns. These variations are primarily attributed to the meteorological modulation of the natural gamma background, as the radionuclide composition of the soil remained constant at the monitoring location throughout the study period. The $\dot{H}^*(10)$ peaked during the summer, while nighttime values were consistently higher than daytime across all seasons. This phenomenon is largely attributed to the nocturnal accumulation of radon decay products within a more stable atmospheric boundary layer. Correlation and regression analyses indicate that temperature and atmospheric humidity indicators are the most stable covariates on a daily scale. Conversely, precipitation events trigger short-term increases in the dose rate lasting several hours. Distinguishing these meteorologically induced fluctuations from potential artificial contamination is essential for robust environmental monitoring. Furthermore, the annual mean $\dot{H}^*(10)$ of 101.32 nSv h^{-1} is consistent with the expected natural background radiation for the regional geological landscape, providing a data-driven confirmation that observed levels remain within normal environmental limits and supporting transparent risk communication regarding historical radiological concerns. Overall, these findings establish local reference values for Vranje and facilitate more accurate data interpretation within the early warning network, which could ultimately reduce the frequency of false alarms during adverse weather events.

References

[1] UNSCEAR: *Sources, Effects and Risks of Ionizing Radiation, UNSCEAR 2024 Report to the General Assembly, with annexes, Volume II, Annex B*, United Nations, New York, 2025.

[2] TCHORZ-TRZECIAKIEWICZ,D.,E., KAMIŃSKA,J.,A.: Variations in gamma radiation and alpha-emitting radionuclides in correlation with weather and location conditions, Scientific Reports, 15(1) (2025) 25063.

[3] NIECKARZ,Z., KUBICKI,M.: I*nfluence of meteorological and cosmic conditions on dose rate in the environment of the Polish Polar station Hornsund*, International Journal of Biometeorology, 69(11) (2025) 2941-2951.

[4] RIZZO,A. et al.: *Environmental Gamma Dose Rate Monitoring and Radon Correlations: Evidence and Potential Applications*, Environments, 9(6) (2022) 66.

[5] JOO,H.,Y., KIM,J.,W., JEONG,S.,Y., KIM,Y.,S., MOON,J.,H.: *Use of big data for estimation of impacts of meteorological variables on environmental radiation dose on Ulleung Island, Republic of Korea*, Nuclear Engineering and Technology, 53(12) (2021) 4189-4200.

[6] DATAR,G., VICHARE,G., SELVARAJ,C., BHASKAR,A., RAGHAV,A.: *Causes of the diurnal variation observed in gamma-ray spectrum using NaI (Tl) detector*, Journal of Atmospheric and Solar-Terrestrial Physics, 207 (2020) 105369.

[7] UNEP: *Depleted Uranium in Serbia and Montenegro: Post-conflict Assessment in the Federal Republic of Yugoslavia*, UNEP, Geneva, 2002.

[8] RTS: U-238 – osiromašeni uranijum, [Online], Available: https://www.rts.rs/lat/tv/rts1/5908274 /u-238--osiromaseni-uranijum.html [Accessed: March. 1, 2026].

[9] LIVESAY,R.,J., BLESSINGER,C.,S., GUZZARDO,T.,F., HAUSLADEN,P.,A.: *Rain-induced increase in background radiation detected by Radiation Portal Monitors*, Journal of Environmental Radioactivity, 137 (2014) 137-141.

[10] EUROPEAN COMMISSION: *Radiological Maps*, [Online], Available: https://remap.jrc.ec. europa.eu/Advanced.aspx [Accessed: Feb. 20, 2026].

[11] ABDUL RASHID,F.,I., IDRIS,M.,I., ABDUL RAHMAN,I., MOHAMED ZIN,M.,R.: *Meteorological effects on atmospheric radioactive particles: a review of recent research*, Journal of Nuclear Science and Technology, 62(10) (2025) 881-914.

[12] REPUBLIC HYDROMETEOROLOGICAL SERVICE OF SERBIA: *Meteorological Annual 1*, 2024, RHSS, Belgrade, 2025.

[13] BACIU,A.,C.: *Outdoor absorbed dose rate in air in relation to airborne natural radioactivity and meteorological conditions at Bucharest (Romania)*, Journal of Radioanalytical and Nuclear Chemistry, 268(1) (2006) 3-14.

[14] MELINTESCU,A., CHAMBERS,S.,D., CRAWFORD,J., WILLIAMS,A.,G., ZORILA,B., GALERIU,D.: *Radon-222 related influence on*

ambient gamma dose, Journal of Environmental Radioactivity, 189 (2018) 67-78.

[15] AMESTOY,J. et al.: *Effects of environmental factors on the monitoring of environmental radioactivity by airborne gamma-ray spectrometry*, Journal of Environmental Radioactivity, 237 (2021) 106695.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INVESTIGATION OF NON-STOICHIOMETRIC TUNGSTEN OXIDES AS SUPPORTS FOR ANODE CATALYSTS IN PEM FUEL CELLS

SNEŽANA BRKOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, snezana.miulovic@vin.bg.ac.rs

IVANA PEROVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, 060 6690400, ivanaperovic@vin.bg.ac.rs

MINA SEOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, mina050@vin.bg.ac.rs

STEFAN MITROVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, stefan.mitrovic@vin.bg.ac.rs

MILENA PIJOVIĆ RADOVANOVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, pijovic@vin.bg.ac.rs

JELENA GEORGIJEVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, georgijevic@vin.bg.ac.rs

PETAR LAUŠEVIĆ
University of Belgrade, Vinča Institute of Nuclear Sciences, National Institute of the Republic of Serbia, Mike Petrovića Alasa 12-14, 11351 Vinča, Serbia, petarl@vinca.rs

Abstract: *The widespread production and application of fuel cells are primarily hindered by the high cost of their components, particularly catalysts. The presence of fuel impurities leads to the degradation and poisoning of catalysts in Proton Exchange Membrane (PEM) fuel cells, significantly reducing their efficiency and lifetime. Consequently, considerable research efforts have been directed toward the development of cost-effective and durable catalysts, as well as suitable support materials. In this study, non-stoichiometric tungsten oxides were synthesized and used as supports for PtRu-based catalysts intended for PEM fuel cells. The electrochemical performance of the obtained PtRu/WO_x+C catalysts was evaluated using cyclic voltammetry and linear sweep voltammetry. The results indicate that the limiting diffusion currents for the hydrogen oxidation reaction are rapidly achieved and remain stable over a wide potential range relevant to anodic operation in PEM fuel cells. The synthesized catalysts exhibit higher catalytic activity and improved resistance to CO poisoning compared to commercial catalysts. Among the investigated samples, the 30% PtRu/WO_x+C catalyst demonstrated the highest catalytic activity and CO tolerance, highlighting its potential as a promising candidate for practical and commercial applications.*

Keywords: *Anode catalyst, CO tolerance, PEM fuel cell.*

1. INTRODUCTION

The development of renewable energy technologies is essential for reducing carbon dioxide and other harmful emissions while improving energy sustainability and security of supply. Among various energy conversion systems, fuel cells have attracted significant attention due to their ability to convert chemical energy directly into electricity and heat. Fuel cell stacks, consisting of multiple individual fuel cells, can be installed near the point of energy consumption, enabling efficient and decentralized energy generation [1,2].

Proton Exchange Membrane (PEM) fuel cells are considered among the most promising technologies due to their high efficiency, low operating temperature, and environmentally friendly operation, producing electricity, heat, and water. However, their large-scale

Corresponding author: S. Brković;
E-mail: snezana.miulovic@vin.bg.ac.rs
DOI: 10.5937/OTEH26115B

commercialization is still limited by high cost and insufficient durability. These challenges are mainly related to catalyst degradation, support oxidation, and sensitivity to fuel impurities [3].

Platinum-group metals, particularly platinum, are widely used as electrocatalysts for hydrogen oxidation and oxygen reduction reactions due to their high activity. However, their high cost and susceptibility to carbon monoxide (CO) poisoning negatively affect the performance and lifetime of PEM fuel cells [4]. Therefore, the development of more durable and cost-effective catalysts and catalyst supports remains a major research focus. The aim of this study is to develop non-stoichiometric tungsten oxide materials as catalyst supports for PtRu anode catalysts in PEM fuel cells. Tungsten oxide is investigated due to its potential to improve CO tolerance and catalyst stability. The influence of different PtRu loadings on electrochemical performance is evaluated using thin-film electrodes combined with rotating disk electrode (RDE) voltammetry.

2. MATERIALS AND METHODS

2.1. Catalyst preparation

In this study, tungsten oxide was synthesized by dispersing 1 g of tungsten powder (Sigma-Aldrich) in 25 cm^3 of deionized water. Subsequently, 25 cm^3 of 30 wt.% hydrogen peroxide (Sigma-Aldrich) was added, followed by the addition of an additional 20 cm^3 of 30 wt.% H_2O_2 after 15 minutes. The suspension was then allowed to evaporate, dried at 110 °C, and a lemon-yellow powder was obtained, designated as WO_x (non-stoichiometric tungsten oxide).

Anode catalysts were prepared by introducing appropriate amounts of commercial hexachloroplatinic acid (H_2PtCl_6, Merck) and ruthenium chloride ($RuCl_3$, Merck) into 35 ml of 0.005 mol dm^{-3} perchloric acid ($HClO_4$, Sigma-Aldrich) containing the synthesized WO_x support. The mixture was subjected to ultrasonic treatment and magnetic stirring for 30 minutes each. Subsequently, 5 ml of 0.1 mol dm^{-3} sodium borohydride solution was added dropwise under continuous stirring, and the suspension was further stirred for 30 minutes. Vulcan XC72R carbon black was then added, followed by an additional 30 minutes of stirring. The resulting suspension was allowed to settle, filtered, washed with deionized water, and dried at 50 °C for 24 hours. The obtained PtRu/WO_x+C catalyst samples contained 10%, 20%, and 30% PtRu loadings.

2.2. Cyclic voltammetry and kinetics

The electrochemical properties of the prepared catalysts were investigated after confirming their composition, using catalyst thin films deposited onto a gold rotating disk electrode (RDE) with a fixed platinum loading of 40 µg cm^{-2}.The synthesized catalysts were evaluated for potential application in PEM fuel cells and compared with a commercial 30% PtRu/C reference catalyst.
Electrochemical measurements were initiated by cyclic voltammetry after purging the 0.5 mol dm^{-3} $HClO_4$ electrolyte with nitrogen for 30 minutes at room temperature. The potential was cycled between 0.03 and 1.2 V vs. RHE at a scan rate of 100 mV s^{-1}. Hydrogen oxidation reaction (HOR) kinetics were investigated using linear sweep voltammetry (LSV) at different rotation rates (1000, 1500, 2000, and 2500 rpm) in the potential range of 0 – 0.3 V vs. RHE. Kinetic and mass activities were calculated using the Koutecky–Levich equation ($1/i = 1/i_k + 1/i_l$) and the specific activity relationship ($i_m = i_k/L$) [5].

The CO tolerance of the catalysts was evaluated by CO adsorption at 50 mV vs. RHE for 2000 s, followed by potential scanning up to 0.9 V and subsequent potential cycling between 0 and 1.2 V vs. RHE to promote CO oxidation.

3. RESULTS AND DISCUSSION

The crystalline structure of the prepared WO_x catalyst support was characterized by X-ray powder diffraction (XRPD) using a Philips PW-1710 diffractometer. The XRPD pattern revealed the coexistence of WO_2 and WO_3 phases, indicating the presence of tungsten in different oxidation states and confirming the non-stoichiometric nature of the WO_x support. The elemental composition of the prepared Pt–Ru catalysts and catalyst support (WO_x) was determined by energy-dispersive X-ray spectroscopy (EDS) using a JEOL JSM-6610LV microscope. The obtained results confirmed the successful preparation of the Pt–Ru catalysts, with a Pt: Ru weight ratio of 2:1 and Pt and Ru mass percentages close to the nominal values, showing only minor deviations.

Figure 1 presents the cyclic voltammograms of the commercial PtRu/C catalyst and the 30% PtRu/WO_x+C catalyst.

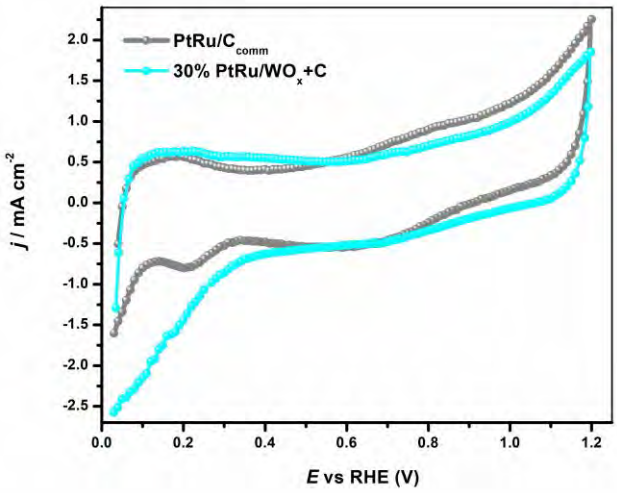


Figure 1. Comparison of the cyclic voltammograms of the commercial and 30% PtRu/WO_x+C catalysts

In the characteristic hydrogen adsorption/desorption region (0.03 V < E < 0.35 V vs. RHE), the PtRu/C catalyst exhibits well-defined and relatively symmetrical anodic and cathodic peaks, which are associated with hydrogen adsorption and desorption processes on the Pt-based active sites.

The H_2 oxidation polarization curves (insets of Figure 2) recorded at different rotation rates for the commercial

30% PtRu/C catalyst (left) and the 30% PtRu/WO_x+C catalyst (right) are presented as insets in Figure 2. For both catalysts, the limiting diffusion current density is reached rapidly, at an overpotential of approximately 60mV, indicating fast hydrogen oxidation kinetics. Furthermore, both catalysts exhibit well-defined and stable polarization behavior over a broad potential range corresponding to the anode operating region of an active PEM fuel cell.

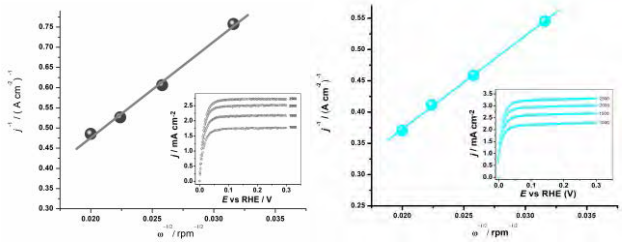


Figure 2. LSV curves of H_2 oxidation at different rotation speeds with the fitted Koutecky-Levich line at 0.02 V vs. RHE for 30% PtRu/C (left) and 30% PtRu/WO_x+C (right)

The current densities at different rotation speeds were extracted from the HOR polarization curves, and the corresponding Koutecky–Levich plots were constructed (Figure 2), for 30% PtRu/C (left) and 30% PtRu/WO_x+C (right). The kinetic current density at an overpotential of 20 mV (j_k) was determined, and the corresponding mass activity (j_m) was calculated for all catalysts (Table 1).

Table 1. Kinetic current densities and mass activities

Catalysts	j_k / *mA cm*$^{-2}$	j_m / *A* mg_{Pt}^{-1}
30% PtRu/C	7,86	0,197
10% PtRu/WO_x-C	10,51	0,263
20% PtRu/WO_x-C	11,40	0,285
30% PtRu/WO_x-C	12,97	0,324

As shown in Table 1, all synthesized catalysts exhibit higher hydrogen oxidation activity than the commercial 30% PtRu/C catalyst. Among the investigated samples, the 30% PtRu/WO_x+C catalyst demonstrated the highest activity, with a kinetic current density of 12.97 mA cm^{-2} and a mass activity of 0.324 A mg_{Pt}^{-1}.

Since the prepared catalysts are intended for application in PEM fuel cells, high tolerance to carbon monoxide poisoning is of crucial importance. Therefore, all synthesized catalysts were evaluated in a gas mixture containing 98% H_2 and 2% CO, simulating the reformate fuel typically used in practical PEM fuel cell systems. The catalysts' tolerance to CO poisoning was evaluated by comparing the HOR polarization curves recorded in pure H_2 and in a gas mixture.

The first cycle of the cyclic voltammogram (CO stripping) corresponds to the electrooxidation of adsorbed CO (CO_{ads}), characterized by a pronounced CO oxidation peak. The second cycle, referred to as the baseline cyclic voltammogram, should coincide with the voltammogram recorded in a pure hydrogen atmosphere, indicating complete removal of the adsorbed CO layer.

The inset of Figure 3 presents the CO_{ads} stripping voltammogram of the commercial 30% PtRu/C catalyst. The first cycle of the cyclic voltammogram, denoted as the CO stripping curve, exhibits a broad and well-defined oxidation peak corresponding to the electrooxidation of CO adsorbed on the catalyst surface. The second cycle, representing the baseline cyclic voltammogram after the removal of adsorbed CO, coincides with the voltammogram recorded in a nitrogen atmosphere. The electrochemically active surface area for CO oxidation (ESA_{CO}) was calculated from the charge associated with the CO oxidation peak (Q, mC) using the equation ESA=Q/q, yielding a value of 0.605 cm².

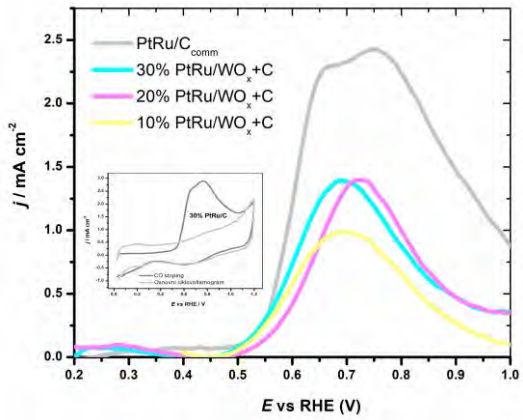


Figure 3. Cyclovoltammogram of CO_{ads} desorption - comparison of CO_{ads} oxidation peaks of commercial, 10%, 20% and 30% PtRu/WO_x+C catalyst,

Figure 3 shows the high-potential region of the cyclic voltammograms, corresponding to the CO_{ads} oxidation peaks of all prepared catalysts. Comparison of the CO stripping voltammograms recorded for the PtRu/WO_x+C catalyst series with that of the commercial PtRu/C catalyst reveals differences over the entire investigated potential range. In the low-potential region (E < 0.5 V vs. RHE), these deviations are attributed to the presence of OH species at the interfacial region between the WOx support and PtRu nanoparticles (PtRu|WO_x [6].

In the high-potential region (E > 0.5 V vs. RHE), all catalysts undergo rapid COads oxidation and removal from the surface. For the commercial PtRu/C catalyst, a broad double oxidation peak is observed, with the main peak located at 0.760 V vs. RHE, whereas the PtRu/WO_x+C catalysts exhibit only a single CO_{ads} oxidation peak. Furthermore, these peaks are shifted by 40–70 mV toward lower potentials compared with the CO_{ads} oxidation peak of the commercial catalyst. In this potential region, CO_{ads} oxidation proceeds mainly through reaction with OH species formed on oxidized metal sites, Pt–OH_{ads} and Ru–OH_{ads}, on both the synthesized catalysts and the commercial PtRu/C catalyst. Since differences in the amount of oxidized CO species are observed between the PtRu/WO_x+C catalysts and the commercial catalyst, an additional effect must contribute to the enhanced CO removal. The improved CO tolerance of the PtRu/WO_x+C catalysts is attributed to the electronic interaction between the WO_x support and PtRu nanoparticles, which modifies the chemisorption properties of the PtRu active site [6]. This electronic effect may also explain the less complex CO oxidation peak profiles observed for the PtRu/WO_x+C catalysts compared with the commercial PtRu/C catalyst.

Table 2. Electrochemical active surface area for CO oxidation and CO oxidation mass activity

Catalysts	ESA_{CO} /cm^2	$j_{m,CO}$ / A mg_{Pt}^{-1}
30% PtRu/C	0.605	0,028
10% PtRu/WO_x-C	0.204	0,050
20% PtRu/WO_x-C	0.270	0,060
30% PtRu/WO_x-C	0.290	0,121

The ESA_{CO} values were calculated from the integrated areas of the CO_{ads} oxidation peaks for the PtRu/WO_x+C catalysts. The obtained values (Table 2) are significantly lower than that of the commercial PtRu/C catalyst.

Figure 4 presents the HOR polarization curves recorded in pure hydrogen and in a 98% H_2 + 2% CO gas mixture at an electrode rotation rate of 2000 rpm and a scan rate of 5 mV s^{-1}. As expected, the anodic current density recorded in the H_2/CO mixture is lower than that obtained in pure hydrogen due to partial blockage of the active Pt sites by adsorbed CO. During the reverse potential scan from 0.9 V to 0 V vs. RHE, the catalyst surface is progressively freed from CO_{ads}, resulting in higher current densities than those observed during the forward scan.

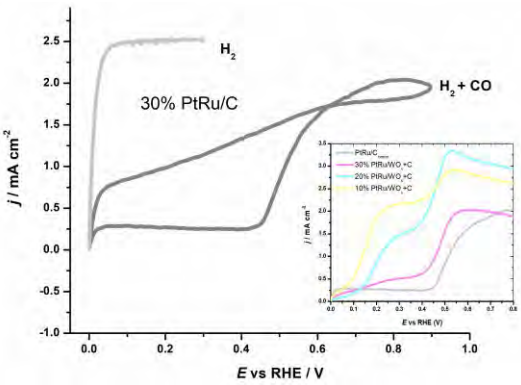


Figure 4. HOR curves for commercial 30% PtRu/C catalyst recorded in pure H_2 and mixture (H_2 +2%CO) gases, at 2000 rpm; inset: HOR curves in H_2 + 2%CO mixture at 2000 rpm - Comparison of anode branches of commercial and PtRu/WO_x+C catalyst

The current density at an overpotential of 20 mV was extracted from the reverse scan, and the corresponding mass activity in the presence of CO ($j_{m,CO}$) was calculated (Table 2). Mass activity $j_{m,CO}$ (A mg_{Pt}^{-1}) in a mixture (98% H_2 + 2% CO) of gases of all analyzed catalysts (inset of Figure 4) was also determined and the obtained values are shown in Table 2.

According to the results summarized in Table 2, all synthesized catalysts outperformed the commercial 30% PtRu/C catalyst in terms of resistance to CO poisoning. Of the investigated samples, the 30% PtRu/WOx+C catalyst achieved the highest mass activity in the presence of CO, demonstrating the best overall electrochemical performance under poisoning conditions.

4. CONCLUSION

WO_x was synthesized and subsequently used as a support for the preparation of Pt- and Ru-containing catalysts. Their electrochemical performance was evaluated by cyclic voltammetry and rotating disk electrode (RDE) voltammetry. Particular emphasis was placed on investigating carbon monoxide poisoning to assess the CO tolerance of the synthesized catalysts.

The results revealed that the mass activities for the hydrogen oxidation reaction (HOR) measured in a gas mixture containing 98% H_2 and 2% CO were considerably lower than those obtained in pure hydrogen. Among the investigated samples, the commercial 30% PtRu/C catalyst exhibited the lowest mass activity, whereas the synthesized PtRu-based catalysts demonstrated significantly improved performance.

The enhanced catalytic activity can be attributed to the formation of the PtRu|support interfacial region, as well as to the generation of Ru–CO_{ads} species, which promote the release of Pt active sites and facilitate the HOR. In addition, most of the synthesized catalysts displayed a shift of the CO_{ads} oxidation/desorption peak toward lower potentials, indicating a higher resistance to CO poisoning compared with the commercial catalyst.

Overall, these findings demonstrate the promising catalytic performance of the synthesized 30% PtRu/WO_x catalyst, particularly in terms of its improved catalytic activity and enhanced CO tolerance, and provide a basis for further evaluation under full-cell conditions.

References

[1] SEBBANI,I,, ETTOUHAMI,M.K., BOULAKHBAR,M: *Fuel cells: A technical, environmental, and economic outlook,* Cleaner Energy Systems, 10 (2025) 100168.

[2] SAHU,S., KANWAL,R., RATNAWAT,I., MIR,A., ABRAR,I.: *Hydrogen fuel cells: technical, economic, and policy pathways toward net-zero integration*, Sustainable Energy Fuels, 9(24) (2025) 6601–6630.

[3] HEGDE,S., WÖRNER,R., SHABANI,B.: *Automotive PEM fuel cell catalyst layer degradation mechanisms and characterisation techniques, Part II: Platinum degradation,* International Journal of Hydrogen Energy, 143 (2025)179-212.

[4] HAFTTANANIAN,M.A., RAMIARA,A., SHABANI,B.B, RANJBAR,A.A.: Nonlinear algorithm of PEM fuel cell catalyst poisoning progress in the presence of carbon monoxide in anode fuel: A computational study using OpenFOAM, Electrochimica Acta, 246 (2017) 348-364.

[5] ARENZ,M., MARKOVIC,N.M.: Half-Cell Investigations of Cathode Catalysts for PEM Fuel Cells: From Model Systems to High-Surface-*Area Catalysts. In: Fuel Cell Sci. Theory*, Fundam. Biocatal., John Wiley & Sons, Inc., New Jersey, 2010.

[6] BRKOVIC,S., NIKOLIC,V., MARCETA KANINSKI,M., PASTI,I.: *Pt/C catalyst impregnated with tungsten-oxide – Hydrogen oxidation reaction vs. CO tolerance,* International Journal of Hydrogen Energy 44 (2019) 13364–13372.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EVALUATION OF B-PINENE AS AN ENVIRONMENTALLY FRIENDLY CORROSION INHIBITOR FOR CARBON STEEL IN 1 M HYDROCHLORIC ACID

ANĐELA R. SIMOVIĆ
Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoseva 12, 11000 Belgrade, Serbia,
andjela.simovic@ihtm.bg.ac.rs

JELENA B. BAJAT
Faculty of Technology and Metallurgy, University of Belgrade, Karnegijeva 4, 11000 Belgrade, Serbia,
jela@tmf.bg.ac.rs

Abstract: *The development of environmentally friendly corrosion inhibitors has attracted increasing attention due to the need for sustainable alternatives to conventional synthetic compounds. In this study, the corrosion inhibition performance of β-pinene, a naturally occurring bicyclic monoterpene commonly found in plant essential oils, was investigated for carbon steel in 1 M hydrochloric acid solution. The inhibitory effect of β-pinene was evaluated using electrochemical impedance spectroscopy (EIS), potentiodynamic polarization measurements, scanning electron microscopy (SEM), and contact angle analysis. The EIS results demonstrated a significant increase in polarization resistance with increasing β-pinene concentration, reaching 292.2 Ω cm² at 60 ppm compared with 23.4 Ω cm² for the uninhibited solution. The inhibition efficiency increased progressively with inhibitor concentration, from 80.0% at 15 ppm to 92.0% at 60 ppm. Potentiodynamic polarization measurements confirmed the concentration-dependent protective effect of β-pinene, with the corrosion current density decreasing from 173.0 μA cm⁻² for the blank solution to 19.2 μA cm⁻² at 60 ppm, corresponding to an inhibition efficiency of 88.9%. The polarization results indicated that β-pinene acts as a mixed-type corrosion inhibitor with a predominant cathodic effect, mainly through adsorption on the carbon steel surface. SEM analysis revealed a considerably smoother and less damaged surface in the presence of β-pinene, confirming the formation of a protective adsorbed layer. Furthermore, contact angle measurements demonstrated an increase in surface hydrophobicity after inhibitor adsorption, supporting the formation of an organic protective film. The obtained results show that β-pinene effectively suppresses carbon steel corrosion in acidic media and highlight its potential as a green corrosion inhibitor and as an active phytochemical contributing to the protective properties of β-pinene-containing essential oils.*

Keywords: *β-pinene; green corrosion inhibitor; carbon steel; electrochemical measurements; SEM.*

1. INTRODUCTION

Metal corrosion in acidic environments remains a major challenge in many industrial processes, including acid pickling, descaling, and cleaning. Hydrochloric acid (HCl) is widely used because of its high efficiency in removing oxide layers and surface contaminants; however, its aggressive nature accelerates metal dissolution. Consequently, corrosion inhibitors are commonly employed to minimize corrosion in acidic media [1,2].

Although conventional corrosion inhibitors are generally effective, many exhibit toxicity and poor biodegradability, which has encouraged the search for environmentally friendly alternatives [3,4]. Among the most promising green corrosion inhibitors are plant-derived essential oils and their phytochemical constituents because of their renewability, biodegradability, and low environmental impact [5,6].

β-Pinene is a naturally occurring bicyclic monoterpene widely distributed in the essential oils of numerous aromatic and medicinal plants. It is frequently found in essential oils that have demonstrated promising corrosion inhibition performance in acidic media [7,8]. However, since essential oils are complex mixtures, their inhibition efficiency cannot be attributed to a single constituent. Therefore, investigating β-pinene as an individual phytochemical is important for understanding its actual contribution to the corrosion protection provided by β-pinene-containing essential oils.

Despite the increasing interest in plant-derived corrosion inhibitors, studies focused on the inhibition behavior of individual phytochemicals remain limited. In particular, the corrosion inhibition performance of β-pinene under well-defined experimental conditions has not been sufficiently investigated.

Therefore, the present study evaluates the corrosion inhibition performance of β-pinene toward carbon steel in 1 M HCl solution. Electrochemical impedance spectroscopy (EIS), potentiodynamic polarization measurements, scanning electron microscopy (SEM), and contact angle measurements were employed to assess its

Corresponding author: A. Simović;
E-mail: andjela.simovic@ihtm.bg.ac.rs

DOI: 10.5937/OTEH26116S

inhibition efficiency and the characteristics of the protective film formed on the steel surface. The obtained results provide further insight into the potential application of β-pinene as an environmentally friendly corrosion inhibitor.

2. EXPERIMENTAL PART

2.1. Materials

β-Pinene (Sigma-Aldrich) was used as the corrosion inhibitor in this study. A stock solution was prepared in 30% (v/v) ethanol, and the ethanol content was maintained at the same level in both the inhibited and blank solutions to eliminate its influence on the corrosion behavior. Test solutions were prepared by dissolving β-pinene in 1 M HCl to obtain concentrations of 15, 30, and 60 ppm.

The corrosion experiments were performed using carbon steel with the following chemical composition (wt.%): Mn 0.276, Cu 0.025, Cr 0.030, Ni 0.012, C 0.080, and Fe as the balance. Before each experiment, the steel specimens were sequentially ground using silicon carbide abrasive papers with grit sizes ranging from 200 to 1500. The polished samples were then rinsed with deionized water, cleaned with absolute ethanol, and dried in air before use.

2.2. Methods

2.2.1. Electrochemical Measurements

The corrosion inhibition performance of β-pinene was evaluated using a GAMRY Reference 600 Potentiostat/Galvanostat/ZRA in a conventional three-electrode electrochemical cell. Carbon steel served as the working electrode, a saturated calomel electrode (SCE) as the reference electrode, and a platinum wire as the counter electrode.

Prior to electrochemical measurements, the working electrode was immersed in the test solution for 1 h to allow stabilization of the open-circuit potential (OCP). Electrochemical impedance spectroscopy (EIS) measurements were carried out at OCP over the frequency range from 100 kHz to 0.01 Hz using a sinusoidal perturbation amplitude of 10 mV and 10 points per decade. Potentiodynamic polarization measurements were subsequently performed within the potential range of ±250 mV versus the corrosion potential at a scan rate of 1 mV s^{-1}. The experimental data were analyzed using Gamry Echem Analyst.

2.2.2. Surface Characterization

The surface morphology of carbon steel specimens after exposure to 1 M HCl in the absence and presence of β-pinene was examined using a JEOL 840A scanning electron microscope. Before immersion, all carbon steel specimens used for SEM observation and contact angle measurements were mechanically polished to a 1000-grit finish to ensure a uniform initial surface.

The wettability of the steel surface was evaluated by contact angle measurements after immersion in the β-pinene-containing solution. Following exposure, the specimens were rinsed with distilled water and dried at room temperature. A droplet of Milli-Q water was placed on the specimen surface, and the contact angle was determined using an optical microscope (Smart 5MP Pro, Delta Optical Instruments, Inc., North Little Rock, AR, USA). The recorded images were analyzed using Image-Pro Plus 4.0 software (Media Cybernetics Inc., Rockville, MD, USA).

3. RESULTS

3.1. Electrochemical Impedance Spectroscopy

The Nyquist plots obtained for carbon steel in 1 M HCl solution in the absence and presence of β-pinene at concentrations of 15, 30, and 60 ppm are presented in Figure 1. In all cases, the impedance spectra consist of a single depressed capacitive semicircle, indicating the presence of one dominant time constant associated with the charge-transfer process at the metal/electrolyte interface. The depressed shape of the semicircle is characteristic of non-ideal capacitive behavior arising from surface heterogeneity, roughness, and the adsorption of inhibitor molecules on the steel surface.

Compared with the uninhibited solution, the diameter of the capacitive semicircle increases markedly in the presence of β-pinene, indicating an increase in polarization resistance and improved corrosion protection. Moreover, the semicircle diameter becomes progressively larger with increasing inhibitor concentration, demonstrating the concentration-dependent inhibition performance of β-pinene. The largest semicircle is observed for the solution containing 60 ppm β-pinene, suggesting the formation of a more compact and protective adsorbed layer on the carbon steel surface [9].

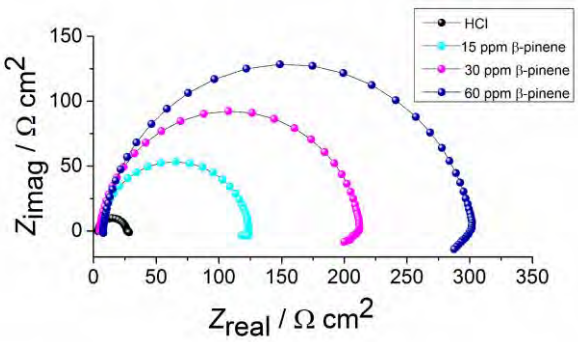


Figure 1. Nyquist plots of carbon steel in 1 M HCl solution in the absence and presence of β-pinene at different concentrations (15, 30, and 60 ppm)

The corresponding Bode plots are presented in Figure 2. The impedance spectra retain the same general shape in both the blank and inhibited solutions, indicating that the addition of β-pinene does not change the corrosion mechanism but rather suppresses the corrosion rate

through adsorption on the metal surface. The phase-angle diagrams exhibit a single maximum in the intermediate-frequency region, confirming the existence of only one time constant and supporting the use of a single-time-constant equivalent electrical circuit for fitting the experimental data.

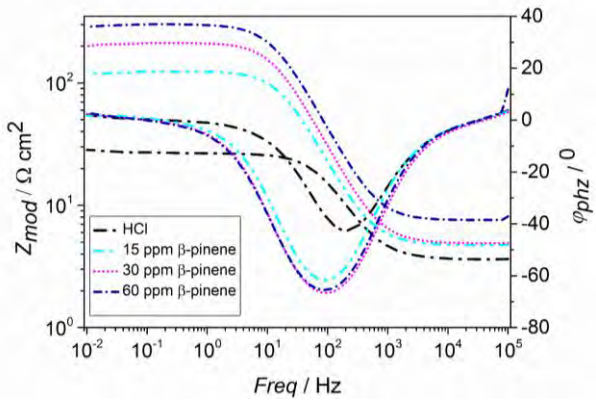


Figure 2. Bode plots of carbon steel in 1 M HCl solution in the absence and presence of β-pinene at different concentrations (15, 30, and 60 ppm)

The equivalent electrical circuit shown in Figure 3 was used to fit the impedance spectra, and the obtained electrochemical parameters are summarized in Table 1. The selected circuit consists of the solution resistance (*Rs*), in parallel with the polarization resistance (*Rp*) and a constant phase element (CPE), which accounts for the non-ideal capacitive behavior of the steel/electrolyte interface caused by surface heterogeneity. The constant phase element (CPE) is characterized by two parameters, Y_o and n, where Y_o represents the CPE magnitude and n is the exponent describing the deviation from ideal capacitive behavior.

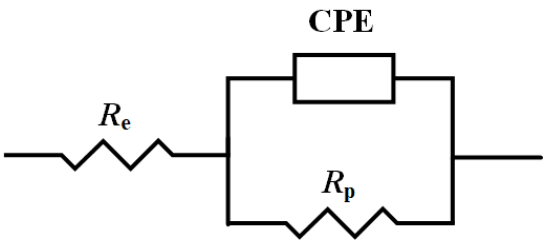


Figure 3. Equivalent electrical circuit used for fitting the EIS data

The inhibition efficiency (η%) was calculated from the polarization resistance values according to the following equation [7]:

$$\eta(\%) = \frac{R_p - R_p^0}{R_p} \times 100 \quad (1)$$

where R_p and R_p^0 are the polarization resistance values obtained in the inhibited and uninhibited solutions, respectively.

As the β-pinene concentration increases, the polarization resistance (*Rp*) increases from 23.4 Ω cm² for the blank solution to 118.6, 205.7, and 292.2 Ω cm² for 15, 30, and 60 ppm, respectively. At the same time, the CPE magnitude (Y_o) decreases from 194.7 × 10^{-6} Ω^{-1} s^n cm^{-2} to 63.4 × 10^{-6} Ω^{-1} s^n cm^{-2}, indicating a reduction in the effective double-layer capacitance due to the adsorption of β-pinene molecules on the steel surface [8]. The inhibition efficiency increases with inhibitor concentration, reaching 80.0%, 88.6%, and 92.0% for 15, 30, and 60 ppm, respectively. These results confirm that β-pinene effectively inhibits carbon steel corrosion in 1 M HCl by forming an adsorbed protective film, with the highest protection achieved at the highest investigated concentration.

Table 1. Electrochemical impedance spectroscopy (EIS) parameters and inhibition efficiency of carbon steel in 1 M HCl solution containing different concentrations of β-pinene.

Solution	R_s / Ω cm²	R_p / Ω cm²	Y_o / 10^{-6}/ Ω^{-1} s^n cm^{-2}	*n*	*η* / %
1 M HCl	3.6	23.4	194.7	0.910	/
15 ppm β-pinene	4.7	118.6	118.2	0,891	80.0
30 ppm β-pinene	4.9	205.7	86.3	0.926	88.6
60 ppm β-pinene	7.4	292.2	63.4	0.923	92.0

3.2. Potentiodynamic Polarization Measurements

The potentiodynamic polarization curves for carbon steel in 1 M HCl solution in the absence and presence of β-pinene at different concentrations are presented in Figure 4. The addition of β-pinene results in a significant decrease in the corrosion current density (jcorr), indicating effective inhibition of the corrosion process. As the inhibitor concentration increases, the corrosion current density decreases progressively from 173.0 μA cm⁻² for the uninhibited solution to 36.8, 27.5, and 19.2 μA cm⁻² in the presence of 15, 30, and 60 ppm β-pinene, respectively.

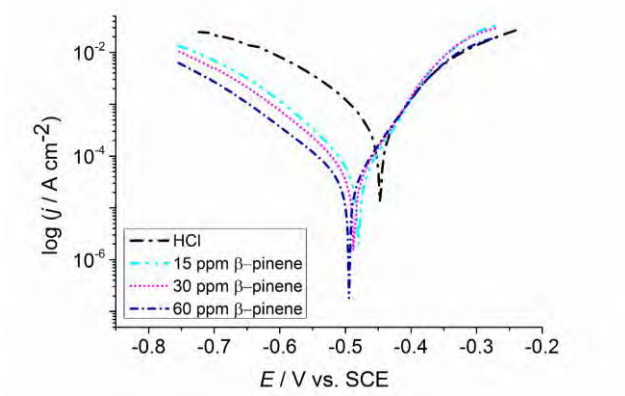


Figure 4. Potentiodynamic polarization curves of carbon steel in 1 M HCl solution in the absence and presence of β-pinene at different concentrations (15, 30, and 60 ppm)

A noticeable reduction in the cathodic branch is observed for all inhibited solutions, whereas the anodic branch remains almost unchanged compared with the blank solution. This behavior indicates that β-pinene predominantly suppresses the cathodic reaction, most likely by blocking the active sites involved in hydrogen evolution through adsorption on the steel surface. The corrosion potential shifts slightly toward more negative values with increasing inhibitor concentration, from −448 mV for the blank solution to −495 mV at 60 ppm β-pinene. Since the displacement of the corrosion potential does not exceed 85 mV, β-pinene may be classified as a mixed-type corrosion inhibitor with a predominant cathodic effect.

The electrochemical parameters obtained by Tafel extrapolation are summarized in Table 2. The inhibition efficiency (η%) was calculated from the corrosion current density values using the following equation [10]:

$$\eta(\%) = \frac{i^0_{corr} - i_{corr}}{i^0_{corr}} \times 100 \quad (2)$$

where i^0_{corr} and i_{corr} represent the corrosion current densities obtained in the absence and presence of β-pinene, respectively.

Increasing the β-pinene concentration leads to a continuous reduction in the corrosion current density, accompanied by an increase in inhibition efficiency from 78.7% at 15 ppm to 88.9% at 60 ppm. These results are consistent with the EIS measurements, confirming that the adsorption of β-pinene on the carbon steel surface effectively suppresses the corrosion process, with the highest protection achieved at the highest investigated concentration.

Table 2. Electrochemical parameters obtained from potentiodynamic polarization measurements for carbon steel in 1 M HCl solution in the absence and presence of β-pinene at different concentrations.

Solution	j_{corr} / μA cm^{-2}	E_{corr} / mV vs. SCE	η / %
1 M HCl	173.0	-448	/
15 ppm β-pinene	36.8	-480	78.7
30 ppm β-pinene	27.5	-488	84.1
60 ppm β-pinene	19.2	-495	88.9

3.3. SEM Analysis

The surface morphology of carbon steel after immersion in 1 M HCl solution in the absence and presence of 60 ppm β-pinene was examined by scanning electron microscopy (SEM), and the corresponding micrographs are presented in Figure 5. The steel specimen exposed to the uninhibited acidic solution (Figure 5a) exhibits extensive surface deterioration characterized by a rough and irregular morphology, confirming the severe corrosive action of hydrochloric acid. The observed damage indicates significant dissolution of the steel surface during exposure to the aggressive medium.

In contrast, the steel sample immersed in the solution containing 60 ppm β-pinene (Figure 5b) displays a considerably smoother and more homogeneous surface with substantially fewer signs of corrosion attack. In several regions, the polishing marks produced during the mechanical preparation of the specimen remain visible, indicating that the steel surface was effectively protected from extensive corrosion. These observations suggest that β-pinene adsorbs onto the carbon steel surface, forming a protective film that acts as a barrier against the aggressive acidic environment. The SEM results are in good agreement with the electrochemical measurements, confirming the effective corrosion inhibition provided by β-pinene.

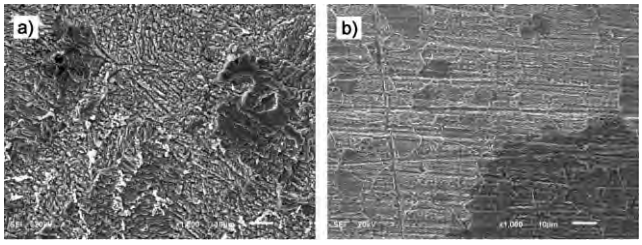


Figure 5. SEM micrographs of carbon steel after immersion in 1 M HCl solution: (a) uninhibited solution and (b) solution containing 60 ppm β-pinene

3.4. Contact Angle Measurements

The wettability of the carbon steel surface after immersion in 1 M HCl solution in the absence and presence of 60 ppm β-pinene was evaluated by contact angle measurements, and the results are presented in Figure 6. The contact angle of the uninhibited steel surface was 37.4°, indicating a highly hydrophilic surface as a consequence of severe corrosion in the acidic medium. In the presence of 60 ppm β-pinene, the contact angle increased to 47.7°, demonstrating a decrease in surface wettability.

The increase in the contact angle suggests that β-pinene adsorbs onto the carbon steel surface, forming a protective organic layer that reduces the affinity of the surface toward water [7]. Although the increase is moderate, it is consistent with the electrochemical and SEM results, confirming the formation of a protective film that limits the interaction between the steel surface and the corrosive environment.

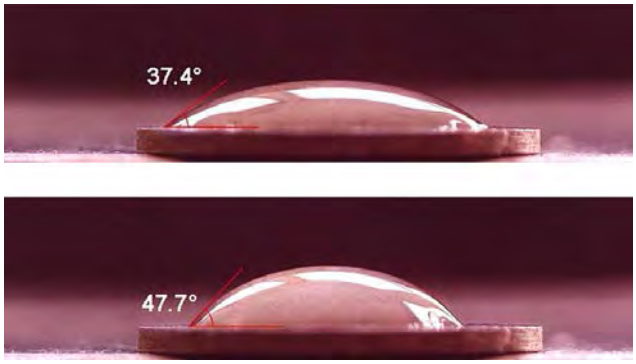


Figure 6. Contact angle measurements of carbon steel after immersion in 1 M HCl solution: without inhibitor (upper image) and in the presence of 60 ppm β-pinene (lower image)

4. CONCLUSION

The present study demonstrated that β-pinene effectively inhibits the corrosion of carbon steel in 1 M HCl, with inhibition efficiency increasing with inhibitor concentration. EIS measurements showed an increase in polarization resistance from 23.4 Ω cm² for the blank solution to 292.2 Ω cm² at 60 ppm β-pinene, corresponding to an inhibition efficiency of 92.0%. Potentiodynamic polarization measurements confirmed the protective effect, with the corrosion current density decreasing from 173.0 to 19.2 μA cm^{-2} and an inhibition efficiency of 88.9% at 60 ppm. The polarization results indicate that β-pinene acts as a mixed-type inhibitor with a predominant cathodic effect. SEM observations revealed substantially lower surface damage in the presence of β-pinene, while the increase in contact angle from 37.4° to 47.7° further supports the formation of a protective organic layer. Overall, the results confirm that β-pinene is a promising environmentally friendly corrosion inhibitor and may represent an important active phytochemical contributing to the protective properties of β-pinene-containing essential oils.

References

[1] Verma, C.; Ebenso, E.E.; Bahadur, I.; Quraishi, M.A. An overview on plant extracts as environmental sustainable and green corrosion inhibitors for metals and alloys in aggressive corrosive media. J. Mol. Liq. 2018, 266, 577–590.

[2] Goyal, M.; Kumar, S.; Bahadur, I.; Verma, C.; Ebenso, E.E. Organic corrosion inhibitors for industrial cleaning of ferrous and non-ferrous metals in acidic solutions: A review. J. Mol. Liq. 2018, 256, 565–573.

[3] Chaubey, N.; Savita; Qurashi, A.; Chauhan, D.S.; Quraishi, M.A. Frontiers and advances in green and sustainable inhibitors for corrosion applications: A critical review. J. Mol. Liq. 2021, 321, 114385.]

[4] Al Jahdaly, B.A.; Maghraby, Y.R.; Ibrahim, A.H.; Shouier, K.R.; Alturki, A.M.; El-Shabasy, R.M. Role of green chemistry in sustainable corrosion inhibition: A review on recent developments. Mater. Today Sustain. 2022, 20, 100242.

[5] Dahmani, K.; Galai, M.; Ouakki, M.; Elgendy, A.; Ez-Zriouli, R.; Lachhab, R.; Briche, S.; Cherkaoui, M. Corrosion inhibition of copper in sulfuric acid via environmentally friendly inhibitor (Myrtus Communis): Combining experimental and theoretical methods. J. Mol. Liq. 2022, 347, 117982.

[6] Damej, M.; Skal, S.; Aslam, J.; Zouarhi, M.; Erramli, H.; Alrashdi, A.A.; Lee, H.S.; El aoufir, Y.; Lgaz, H. An environmentally friendly formulation based on Cannabis sativa L. seed oil for corrosion inhibition of E24 steel in HCl medium: Experimental and theoretical study. Colloids Surf. A Physicochem. Eng. Asp. 2022, 643, 128745.

[7] Simović, A.R.; Novaković, J.; Milovanović, B.; Etinski, M.; Janaćković, P.; Bajat, J.B. Nature-Derived corrosion inhibition of steel in HCl by Pančić spruce essential Oil: Insights from experimental and Computational Approaches. Appl. Surf. Sci. 2025, 698, 163048.

[8] Simović, A.R.; Grgur, B.N.; Novaković, J.; Janaćković, P.; Bajat, J. Black Pine (Pinus nigra) Essential Oil as a Green Corrosion Inhibitor for Carbon Steel. Metals 2023, 13, 508.

[9] Simović, A.; Stevanović, S.; Milovanović, B.; Etinski, M.; Bajat, J.B. Green corrosion inhibitors of steel based on peptides and their constituents: A combination of experimental and computational approach. J. Solid State Electrochem. 2023, 27, 1821–1834.

[10] Qiang, Y.; Zhang, S.; Tan, B.; Chen, S. Evaluation of Ginkgo leaf extract as an eco-friendly corrosion inhibitor of X70 steel in HCl solution. Corros. Sci. 2018, 133, 6–16.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

FRICTION STIR WELDING IN THE JOINING OF LIGHT METAL ALLOYS: A REVIEW

NADA RATKOVIC
University of Kragujevac, Faculty of Engineering, Kragujevac, nratkovic@kg.ac.rs

VESNA MANDIC
University of Kragujevac, Faculty of Engineering, Kragujevac, mandic@kg.ac.rs

ZIVANA JOVANOVIC PESIC
University of Kragujevac, Faculty of Engineering, Kragujevac, zixi90@gmail.com

MARKO DELIC
University of Kragujevac, Faculty of Engineering, Kragujevac, marko.delic@kg.ac.rs

DJORDJE IVKOVIC
University of Kragujevac, Faculty of Engineering, Kragujevac, djordje.ivkovic@fink.rs

Abstract: *Friction Stir Welding (FSW) and Friction Stir Spot Welding (FSSW) are solid-state joining processes in which bonding occurs below the melting temperature of base materials. These techniques are increasingly applied to non-ferrous metals such as aluminum, titanium, magnesium, and copper, as well as to dissimilar material combinations like Al-Ti and Al-Mg. Their growing use is driven by industrial demands, particularly in the aerospace and automotive sectors, where high strength and low weight are essential. This paper presents an overview of conventional and advanced FSW techniques, focusing on the mechanical and microstructural properties of the resulting joints. Experimental studies and industrial applications demonstrate that high-quality joints in aluminum and titanium alloys can be achieved through proper selection of process parameters and tool geometry. Due to the complex thermomechanical and metallurgical processes involved, careful control of welding parameters such as tool rotational speed, welding speed, and dwell time is crucial. These parameters significantly influence joint quality and the formation of intermetallic compounds, such as Ti_3Al. FSW continues to attract research interest because of its efficiency and potential in joining light metal alloys. Additionally, it is an energy-efficient and environmentally friendly process, contributing to sustainable engineering practices.*

Keywords: *Friction Stir Welding (FSW), Friction Stir Spot Welding (FSSW), aluminum, titanium.*

1. INTRODUCTION

Friction Stir Welding (FSW) is an advanced solid-state joining technology that has gained increasing importance across a wide range of industrial applications. This process enables the reliable joining of materials while providing superior joint properties in both similar and dissimilar metal combinations. FSW was invented and patented by The Welding Institute (TWI), United Kingdom, in 1991, and was initially developed as an alternative to conventional fusion welding for aluminum alloys. The technology was rapidly adopted in the aerospace industry, including NASA space programs, where weight reduction and high structural integrity are of primary importance. Over time, the application of FSW has expanded beyond aluminum alloys to include titanium, magnesium, copper, nickel, lead, zinc, and even steels, demonstrating its versatility for joining a broad range of engineering materials [1].

The development and industrial implementation of FSW have been predominantly associated with the welding of aluminum and its alloys, accounting for approximately 99% of its applications. Butt joints represent the majority of FSW joints because of their excellent mechanical properties, as demonstrated by numerous studies [2]. The weldability of aluminum alloys by conventional fusion welding processes is often limited; however, FSW has provided an effective solution for joining these alloys and related materials. Aluminum alloys, particularly those of the 2xxx, 5xxx, 6xxx, and 7xxx series, are widely used in various industrial sectors, including the automotive, aerospace, space, shipbuilding, construction, bridge engineering, and military industries. Their widespread application is primarily attributed to their low density, favorable mechanical properties, and excellent corrosion resistance [3].

The increasing demand for advanced engineering structures has created a growing need for joining both similar and dissimilar materials. In particular, combinations such as Al-Ti, Al-Mg, Cu-Al, and Al-steel have attracted considerable attention due to their potential for achieving an optimal balance between mechanical

Corresponding author: Z. Jovanovic Pesic;
E-mail: zixi90@gmail.com
DOI: 10.5937/OTEH26117R

performance, weight reduction, and functional properties. However, joining materials with significantly different melting temperatures by conventional fusion welding processes remains a major challenge. High-strength aluminum and titanium alloys, which are widely used in aerospace applications, are particularly difficult to join using fusion welding because of the metallurgical incompatibilities and defects associated with melting-based processes.

The continued development of FSW has led to significant advances in a number of process variants, including multi-pass FSW, multi-tool FSW, tandem friction stir welding, and Friction Stir Spot Welding (FSSW). These developments have expanded the capabilities of the process and improved its applicability to a wider range of engineering materials and joint configurations. This paper provides an overview of these FSW techniques and discusses their main characteristics. Their implementation has resulted in improved joint performance and significantly broadened the range of compatible alloy systems that can be successfully joined using friction stir welding.

2. FUNDAMENTALS OF THE FSW PROCESS

Friction Stir Welding (FSW) differs from conventional friction welding in that it employs a non-consumable rotating tool to produce the joint. For this reason, FSW is often referred to as tool-assisted friction welding. Unlike conventional fusion welding processes, FSW is a solid-state joining process in which the base materials are joined without melting. During the process, the rotating tool generates frictional heat and severe plastic deformation, softening the material around the tool and enabling its plastic flow. The tool performs a combined rotational and translational motion (Fig. 1). It rotates about its own axis while simultaneously traversing along the joint line, which may be either straight or curved, depending on the joint geometry.

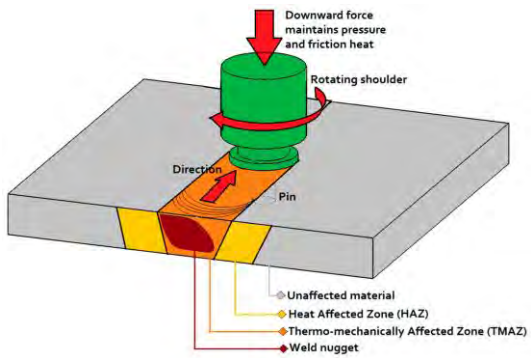


Figure 1. Schematic illustration of the Friction Stir Welding (FSW) process [4]

During the welding process, frictional heat is generated, raising the temperature of the material to approximately 0.8 of its melting temperature, which is sufficient to promote plastic deformation without melting. For most aluminum alloys, the maximum process temperature does not exceed 450 °C [5]. The FSW tool consists of two main components: a shoulder and a pin (Fig. 2). The threaded pin promotes continuous stirring and plastic flow of the softened material around the rotating tool, while the rotating tool typically operates at rotational speeds ranging from 100 to 1500 rpm.

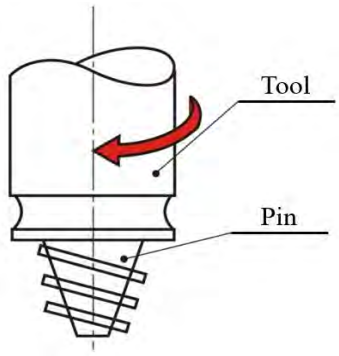


Figure 2. Schematic illustration of the FSW tool [6]

As the tool traverses along the joint line, the softened material undergoes intense plastic deformation and is transported from the leading edge to the trailing edge of the tool. Upon cooling, the plastically deformed material consolidates, forming a sound solid-state joint. At the beginning of the welding process, the tool pin penetrates the workpieces until the tool shoulder comes into contact with the upper surface of the material. After tool withdrawal, a characteristic exit hole, commonly referred to as the keyhole, remains at the end of the weld. The primary function of the shoulder is to apply forging pressure, contain the softened material within the welding zone, and generate the majority of the frictional heat. Tools with profiled shoulder surfaces, such as concentric grooves, spiral features, or scrolls, enhance heat generation and material flow. These surface features facilitate the transport of softened material around the tool, thereby improving material flow, joint formation, and the overall weld quality.

3. FSW-BASED TECHNIQUES

Over a relatively short period, the development of Friction Stir Welding (FSW) has led to the emergence of a broad family of friction stir-based technologies that continue to evolve. These developments have expanded the application of FSW not only as a joining technology but also as a material processing technique. The continuous development of new process variants has expanded the application of FSW to increasingly complex joint geometries and diverse engineering applications. In addition, advances in tool design have contributed to higher process efficiency, improved joint quality, and increased productivity.

The most widely used FSW-based technologies include:

- ✓ Friction Stir Lap Welding (FSLW);
- ✓ Friction Stir Processing (FSP);
- ✓ Friction Stir Dovetailing (FSD);
- ✓ Friction Stir Spot Welding (FSSW);
- ✓ Refill Friction Stir Spot Welding (Refill FSSW);
- ✓ Stationary Shoulder Friction Stir Welding (SSFSW);
- ✓ Underwater Friction Stir Welding (UFSW);
- ✓ Friction Stir Deposition (FSD),

as well as numerous other variants of friction stir-based processing and joining technologies.

Friction Stir Lap Welding (FSLW) is a solid-state joining process primarily used for lap joints in lightweight materials, including aluminum and magnesium alloys, as well as certain polymers [7].

Friction Stir Processing (FSP) is a versatile solid-state processing technique used to refine the microstructure and improve the mechanical properties of metallic materials. It is widely employed in the fabrication of surface composites and has also been successfully combined with ultrasonic vibration-assisted processing to enhance alloying and material modification [8].

Friction Stir Dovetailing (FSD) is a recently developed joining technique for dissimilar materials, particularly aluminum and steel. The incorporation of an aluminum alloy interlayer (AA6061) has been shown to improve the mechanical performance of the joint by increasing both load-bearing capacity and ductility, thereby effectively suppressing the formation of brittle intermetallic compounds [8].

Friction Stir Spot Welding (FSSW) (Fig. 3) is a solid-state spot joining process developed as an alternative to resistance spot welding. The technology is capable of producing high-quality spot welds through precise control of the process parameters and provides excellent mechanical performance, particularly in terms of joint strength. Consequently, FSSW has found widespread application in the automotive and aerospace industries [9].

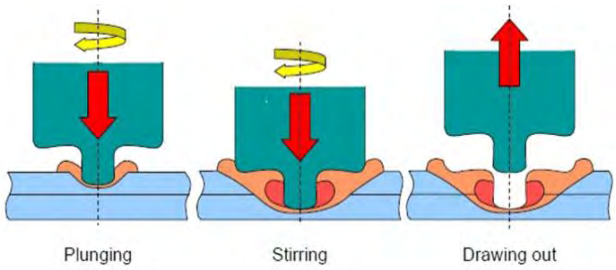


Figure 3. Schematic illustration of the Friction Stir Spot Welding (FSSW) process [10]

Stationary Shoulder Friction Stir Welding (SSFSW) has been shown to provide improved mechanical properties and a more refined microstructure compared with conventional FSW. Owing to its ability to reduce heat input while maintaining excellent joint quality, SSFSW is considered a promising technique for manufacturing lightweight structures [8].

Underwater Friction Stir Welding (UFSW) is an emerging variant of FSW in which the welding process is performed in a water environment. The presence of water creates unique thermal boundary conditions that promote rapid heat dissipation, resulting in a finer microstructure and enhanced joint strength [11]. Owing to these advantages, UFSW has recived increasing attention for joining materials that are difficult to weld using conventional fusion welding processes.

The continuous development of FSW has been driven not only by the need to improve joint quality but also by the demand for higher process efficiency, increased productivity, reduced energy consumption, and greater process flexibility. Consequently, several innovative process configurations have been developed, including multi-tool and double-sided FSW techniques.

Double-sided FSW employs two tools positioned on opposite sides of the workpiece with a slight offset in the welding direction (Fig. 4). Simultaneous counter-rotating operation offers several advantages, including reduced reaction torque and improved weld symmetry. Although the tool pins do not necessarily come into contact, they are positioned sufficiently close to promote effective material flow and mixing throughout the thickness of the joint, ensuring complete weld formation. This configuration also minimizes the asymmetry of lateral forces, resulting in improved process stability and tool balance. Double-sided FSW has been successfully applied to the welding of plates with thicknesses of up to 100 mm.

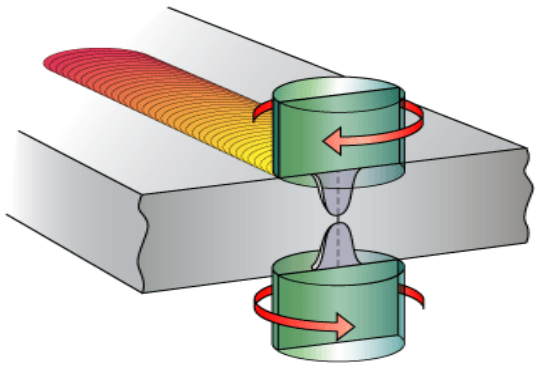


Figure 4. Schematic illustration of the double-sided FSW process [12]

Multi-tool Friction Stir Welding (FSW) enables the simultaneous joining of multiple plates, thereby increasing process productivity and efficiency (Fig. 5).

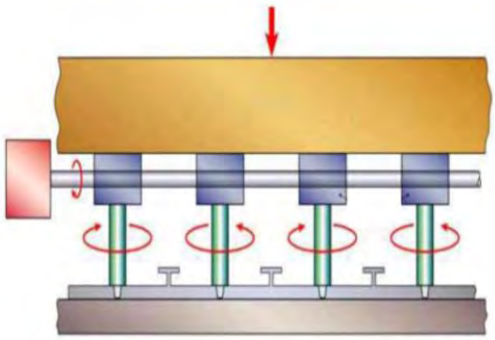


Figure 5. Schematic illustration of the multi-tool FSW process [13]

Several dual-tool FSW configurations have been developed to improve productivity, weld quality, and process efficiency. The three most common configurations are:

1. Parallel twin-tool FSW;
2. Tandem FSW;
3. Staggered twin-tool FSW.

The parallel twin-tool FSW configuration employs two tools that travel simultaneously along parallel welding

paths. This arrangement enables higher welding speeds, threby improved productivity by reducing the overall welding time. Figure 6 illustrates the schematic configuration of this process applied to a lap joint.

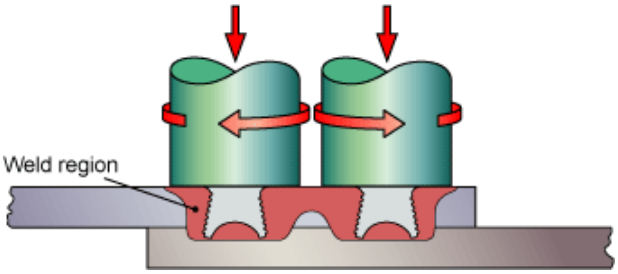


Figure 6. Schematic illustration of the parallel twin-tool FSW configuration [13]

Tandem FSW employs two or more tools positioned sequentially along the welding direction (Fig. 7). The leading tool generates the initial heat input and performs the primary plasticization and stirring of the material while overcoming most of the mechanical resistance. The trailing tool subsequently traverses the softened material, promoting additional material mixing, disrupting residual oxide layers, and refining the weld microstructure. In tandem FSW, the tools are commonly counter-rotating to reduce reaction torque and improve process stability.

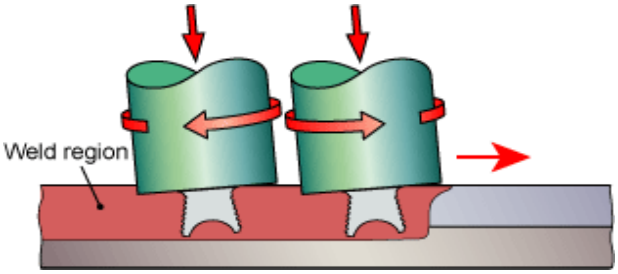


Figure 7. Schematic illustration of the tandem FSW configuration [13]

Tandem FSW has found widespread application in high-performance industries, including aerospace, automotive, and railway engineering. Typical applications include the fabrication of aircraft fuselage panels, rocket propellant tanks, battery enclosures for electric vehicles, and large extruded aluminum profiles used in high-speed railway vehicles [13]. The process is primarily employed for joining aluminum, magnesium, and copper alloys. In shipbuilding, for example, tandem FSW is used for welding thick aluminum alloy plates, particularly those of the 5xxx, 6xxx, and 7xxx series, with thicknesses exceeding 20 mm.

The staggered twin-tool FSW configuration utilizes two tools positioned sequentially with a lateral offset, allowing the second tool to partially overlap the weld produced by the first tool (Fig. 8). This arrangement creates a wider weld region, making it particularly suitable for lap joints where higher joint strength is required. Furthermore, the second tool promotes additional fragmentation and dispersion of residual oxide layers while improving material flow within the thermo-mechanically affected region. Another important advantage of this configuration is its ability to compensate for local thickness reduction that may occur during the first welding pass.

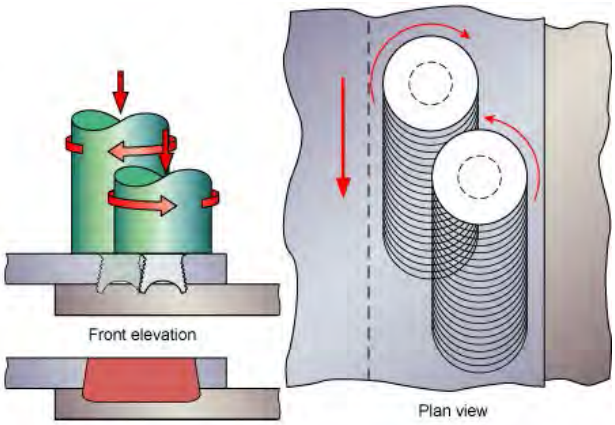


Figure 8. Schematic illustration of the staggered twin-tool FSW configuration [13]

4. WELDABLE MATERIALS

In principle, Friction Stir Welding (FSW) can be applied to joining of nearly all engineering metals, ranging from low-melting-point materials such as aluminum and magnesium to high-melting-point metals including titanium, copper, nickel, and steels. Successful welding depends primarily on the appropriate selection of process parameters and tool geometry, which largely determines the versatility and broad industrial applicability of the process.

Aluminum and its alloys are by far the most extensively welded materials using FSW. All major aluminum alloy series can be successfully joined, making aluminum the primary material group for which FSW was originally developed and remains the most widely adopted in industrial applications.

Magnesium alloys are also highly suitable for FSW. Aluminum-zinc alloys of the AZ series, particularly AZ31 and AZ91, exhibit excellent weldability and are widely used in the automotive and aerospace industries. Similarly, zinc-zirconium alloys of the ZK series produce high-quality welds. FSW has been successfully applied to magnesium components ranging from thin sheets with a thickness of approximately 1 mm to thick plates exceeding 20 mm.

FSW of titanium and its alloys offers an effective joining solution for aerospace, biomedical, and defense applications. Although FSW generally provides superior joint quality compared with conventional TIG and, in some cases, even laser welding, its widespread industrial implementation remains limited because of severe tool wear and the high cost of tool materials. FSW tools for titanium welding are typically manufactured from tungsten-based alloys, tungsten carbide, or polycrystalline cubic boron nitride (PCBN) due to the demanding processing conditions [14].

Compared with other non-ferrous metals, titanium presents unique challenges for FSW because of its high melting temperature and relatively low thermal conductivity. Consequently, lower tool rotational and

welding speeds are generally required to control heat input and minimize tool wear. In the study reported in [15], the mechanical properties and microstructure of the Ti-6Al-4V alloy were investigated using scanning electron microscopy (SEM) and electron backscatter diffraction (EBSD). Microstructural analysis of specimens produced under different processing conditions revealed significant grain refinement and a homogeneous microstructure within the stir zone.

Owing to its excellent biocompatibility, titanium is one of the most important materials for biomedical implants. Consequently, various FSW-based techniques have been developed for fabrication and joining titanium components intended for medical applications.

5. MICROSTRUCTURE OF FSW JOINTS

Owing to the unique characteristics of the FSW process, which combines frictional heating with severe plastic deformation, FSW joints exhibit an asymmetric and highly heterogeneous microstructure. A typical FSW joint consists of four distinct regions: the Base Material (BM), the Heat-Affected Zone (HAZ), the Thermo-Mechanically Affected Zone (TMAZ), and the Stir Zone (SZ), also referred to as the Weld Nugget (WN) (Fig. 9).

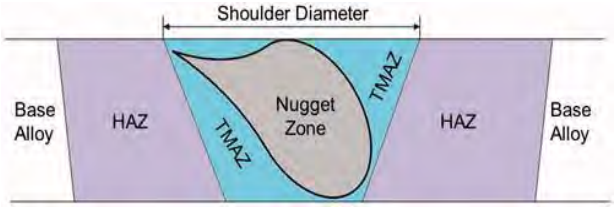


Figure 9. Schematic illustration of the microstructural regions in an FSW joint [16]

The extent of the Heat-Affected Zone (HAZ) depends primarily on the thermal cycle imposed on by the material, particularly the peak temperature and exposure time. In the Thermo-Mechanically Affected Zone (TMAZ), the material undergoes significant plastic deformation combined with elevated temperatures; however, complete recrystallization generally does not occur. The Stir Zone (SZ), located at the center of the weld, experiences the highest levels of plastic deformation and frictional heating. Consequently, extensive dynamic recrystallization occurs within this region, producing a fine equiaxed grain structure.

The boundary between the Stir Zone and the surrounding TMAZ is usually well defined in aluminum alloys because of the distinct difference in recrystallization behavior. In contrast, many other metallic materials exhibit a more gradual transition between these regions. For example, in Ti-6Al-4V and commercially pure copper, the boundaries between the HAZ, TMAZ, and SZ are often less pronounced owing to their different thermomechanical responses during welding.

The Stir Zone is characterized by a fine-grained microstructure, with grain sizes typically ranging from 2 to 10 µm, depending on the alloy system and welding conditions [17]. Figure 10 presents the microstructure of an FSW joint produced in AlMg2.5Cr0.2 aluminum alloy with a plate thickness of 2 mm, illustrating the characteristic microstructural regions.

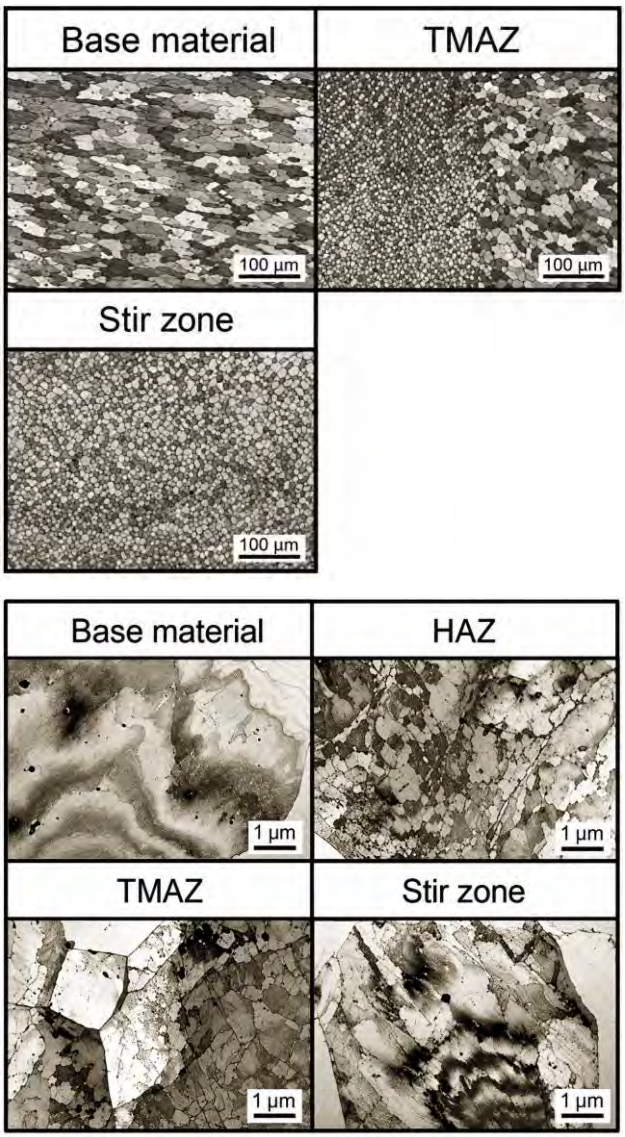


Figure 10. Microstructure of an FSW joint in AlMg2.5Cr0.2 aluminum alloy showing the characteristic microstructural regions [17]

The microstructural evolution during FSW depends strongly on the material being welded and is governed by the combined effects of severe plastic deformation and frictional heating. Dynamic recrystallization occurs within the Stir Zone, leading to the transformation of the original coarse grains into fine equiaxed grains. In aluminum alloys, this process is predominantly associated with geometric dynamic recrystallization. Outside the Stir Zone, where both the deformation and temperature are lower, the grains generally retain their elongated morphology, while dislocations rearrange to form subgrain structures.

Different strengthening mechanisms influence the microstructural evolution of aluminum alloys. In precipitation-hardened alloys, such as AA5052, AA6061, and Al-Mg-Si alloys, the thermal cycle during FSW promotes precipitate dissolution within the Stir Zone, followed by precipitate coarsening during cooling. This

phenomenon is considered one of the primary causes of the localized reduction in hardness. In magnesium alloys, which possess a hexagonal close-packed (HCP) crystal structure with a limited number of slip systems, elevated temperatures activate additional slip systems and deformation twinning. Consequently, dynamic recrystallization becomes more pronounced, often resulting in improved mechanical properties, particularly near the interface between the Stir Zone and the TMAZ [18].

During the FSW of dissimilar light-metal alloys, intense plastic deformation and material flow produce a characteristic swirl-like microstructure. The principal metallurgical challenge in such joints is interfacial diffusion, which may promote the formation of brittle intermetallic compounds. For example, Al-Mg joints typically contain a bilayer intermetallic region composed of Al_3Mg_2, formed adjacent to the aluminum side, and $Al_{12}Mg_{17}$, formed adjacent to the magnesium side. Similarly, a thin $TiAl_3$ intermetallic layer commonly develops at the interface of Al-Ti joints. Therefore, careful control of welding parameters, particularly welding speed and tool geometry, is essential for minimizing the thickness of these intermetallic layers and preserving joint integrity [18].

6. CONCLUSION

Over the past three decades, Friction Stir Welding (FSW) has evolved into one of the most advanced solid-state joining technologies owing to its ability to produce high-quality joints while offering high energy efficiency and excellent environmental sustainability. Nevertheless, further research is required to expand its application to a broader range of engineering materials, complex joint geometries, and dissimilar material combinations.

Although considerable progress has been achieved, several aspects of FSW and Friction Stir Spot Welding (FSSW) remain insufficiently understood. In particular, the joining of dissimilar materials, such as aluminum and titanium, continues to present significant scientific and technological challenges. A deeper understanding of material flow, interfacial reactions, and the formation of intermetallic compounds is essential for further improving joint quality and process reliability.

For lightweight alloys, FSW has become not only an attractive alternative to conventional fusion welding but, in many applications, the only technically feasible and economically viable joining process capable of producing high-quality solid-state joints. Future developments should therefore focus on improving tool materials and extending tool life, particularly when welding high-strength and high-melting-point alloys, where severe tool wear remains one of the primary limitations to large-scale industrial implementation.

The microstructure developed during FSW plays a decisive role in determining the final mechanical performance of the joint. Dynamic recrystallization within the Stir Zone (SZ) produces a refined equiaxed grain structure, resulting in enhanced fatigue resistance, increased hardness, and improved overall mechanical properties.

Further advances in FSW technology will rely on the integration of automation, robotics, digital manufacturing, and numerical simulation. Within the framework of smart manufacturing, virtual process simulation and digital optimization of welding parameters prior to production are becoming increasingly important for improving process efficiency, weld quality, and process repeatability. Consequently, future research should increasingly focus on the development of intelligent process monitoring and control systems to support the wider industrial adoption of FSW technologies.

Finally, one of the most significant advantages of FSW is its environmental sustainability. Since the process is performed entirely in the solid state, it does not require filler materials or shielding gases and produces no harmful fumes, spatter, or intense radiation typically associated with conventional fusion welding. These characteristics make FSW an environmentally friendly and resource-efficient manufacturing process that is widely recognized as a green technology with significant potential for future sustainable industrial applications.

References

[1] RUDRA,P.S., SOMIL,D., AMAN,S., SUBODH,K.: *A review paper on friction stir welding process*, *Materials Today: Proceedings*, 38(1) (2021) 6–11.

[2] KRASNOVSKI,K.: *Fatigue and static properties of friction stir welded aluminium alloy 6082 lap joints using Triflute-type and smooth tools*, *Metallurgy and Materials*, 59 (2014). DOI: 10.2478/amm-2014-0025.

[3] NEVES,M., IVAN,G., RUI,M.L., JOSE,D.C., ALTINO,L.: *Nugget formation and mechanical behaviour of friction stir welding of three dissimilar aluminium alloys*, *Materials*, 13(11) (2020) 2664.

[4] TAHERI,H., KILPATRICK,M., NORVALLS,M., HARPER,W.J., KOESTER,L.W., BIGELOW,T., BOND,L.J.: *Investigation of nondestructive testing methods for friction stir welding*, *Metals*, 9(6) (2019) 624. DOI: 10.3390/met9060624.

[5] DIMOV,N.: *Dissimilar Al6061 to Al7075 Friction Stir Welds: Processing, Static and Fatigue Properties*, Ph.D. Dissertation, Institut Polytechnique de Paris; Université catholique de Louvain, 2022.

[6] RATKOVIĆ,N.: *Modeliranje procesa zavarivanja trenjem mašinskih delova različitih oblika i materijala*, Ph.D. Dissertation, Faculty of Mechanical Engineering, University of Kragujevac, Kragujevac, Serbia, 2009.

[7] GHIYA,R., BADHEKA,V.J.: *A review of friction stir lap welding of polymer to metal*, *Polymer-Plastics Technology and Materials*, 60(18) (2021) 1966–1995. DOI: 10.1080/25740881.2021.1937646.

[8] SEN,M., OZCAN,M.E., YILDIZ,Y.O., AVER,M., KAPAN,S., HUSEYINOGLU,M., OSMAN,Y., SERTAC,E.K.: *A comprehensive study on Friction Stir Welding: A review*, *Soldagem & Inspeção*, 30

(2025) e3016.

[9] SEERANGAN,R.N., CHINNASAMY,R., TUSHAR,S., MIKHAIL,I., KALIYAPERUMAL,B., HARI,B.N., SUBBIAH,S., CLEMENT,V.K.: *Optimization of CNC-FSSW parameters for dissimilar spot welding of AA6061 aluminium alloy and mild steel using Taguchi-based desirability function approach*, *International Journal on Interactive Design and Manufacturing*, 19 (2025) 337–346.

[10] REZAZADEH,N., MOSAVIZADEH,S.M., AZIZI,H.: *Investigation of heat transfer and material flow of P-FSSW: Experimental and numerical study*, *Heat and Mass Transfer*, 54(9) (2018) 2651–2659. DOI: 10.1007/s00231-017-2273-3.

[11] TALEBIZADEHSARDARI,P., MUSHARAVATI,F., KHAN,A., SEBDEY,T.A., EYVAZIAN,A., AGHAJANI DERAZKOLA,H.: *Underwater friction stir welding of Al-Mg alloy: Thermo-mechanical modeling and validation*, *Materials Today Communications*, 26 (2021) 101965. DOI: 10.1016/j.mtcomm.2020.101965.

[12] AFSHARI,M., FAKHRALMOBASHERI,N., SAMADI,M.R., ALAVI,A., FOROUSHANI,H.N.: *Mechanical and formability evaluation of ST14 alloys welded by friction stir welding*, *Shock and Vibration*, 2021(1) (2021) 6269135. DOI: 10.1155/2021/6269135.

[13] THOMAS,W.M., NORRIS,M.I., STAINES,D.G., WATTS,E.R.: *Friction Stir Welding—Process Developments and Variant Techniques*, Proceedings of the SME Summit 2005, Oconomowoc, Milwaukee, USA, TWI Ltd., August 2005.

[14] RUSSELL,M.J.: *FSW of Titanium Alloys—A Progress Update*, Proceedings of the 10th World Conference on Titanium, Hamburg, Germany, July 2003, TWI Ltd., Cambridge.

[15] JAO,J., CHEN,L., DAI,G., GUO,J., LI,W., CHANG,H.: *Microstructure evolution and mechanical property enhancement of friction stir additive manufactured Ti-6Al-4V alloy*, *Journal of Alloys and Compounds*, 987 (2024) 174147.

[16] ARYA,P.K., JAIN,N.K., MURUGESAN,J., PATEL,V.K.: *Developments in friction stir welding of aluminium to magnesium alloy*, *Journal of Adhesion Science and Technology*, 36(13) (2022) 1365–1402. DOI: 10.1080/01694243.2021.1975614.

[17] RAGAB,M., AHMED,M.M.Z., EL-SAYED SELEMAN,M.M., ATAYA,S., ALAMRY,A., EL-SAYED,T.A.: *Friction Stir Welding: A Critical Review of Analytical, Numerical, and Experimental Methods for Quantifying Heat Generation*, *Machines*, 14(4) (2026) 440. DOI: 10.3390/machines14040440.

[18] SHAH,L.H.A., SHAMSOLHODAEI,A., WALBRIDGE,S., GERLICH,A.: *Microstructure and Texture Evolution of Friction-Stir-Welded AA5052 and AA6061 Aluminum Alloys*, *Metals*, 16(1) (2026) 73. DOI: 10.3390/met16010073.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

GRAPHENE OXIDE FOR ORGANOPHOSPHATE REMOVAL – A DFT INSIGHT

DRAGANA VASIĆ ANIĆIJEVIĆ
Department of Physical Chemistry, "VINČA" Institute of Nuclear Sciences – National Institute of the Republic of Serbia, University of Belgrade, Belgrade, dragana.vasic87@gmail.com

VLADAN J. ANIĆIJEVIĆ
Military Academy, University of Defence in Belgrade, Belgrade, anicijevicj.v@gmail.com

Abstract: *Organophosphate removal is crucial for protection of human health, preservation of biodiversity, and prevention of long-term contamination of soil and water systems. Organophosphate removal using carbon-based materials is widely investigated and employed because these materials are widely available, low-cost, and simple to use, making them practical for treating contamination under diverse conditions. The density functional theory (DFT) is a common tool of computational chemistry to assess the molecular aspects of the interaction between organic molecules and ordered supports such as graphene and modified graphene, including: adsorption geometry, adsorption energetics, nature of interaction (chemisorption of physisorption), and mechanistic insights. On the other hand, such task becomes more complex when the adsorbate and the support are rich in functional groups and diverse in steric properties. This contribution deals with the DFT investigation of adsorption of several selected and commonly used organophosphate model molecules on graphene oxide. The goal was to find the optimal binding geometry among several selected input positions and to find energetically most stable orientations. The obtained output geometries were further analyzed and discussed in order to design more precise carbon-based materials for advanced purposes, such as reusable adsorbents or catalysts for irreversible organophosphate hydrolysis.*

Keywords: *density functional theory, adsorption, organophosphate, graphene oxide.*

1. INTRODUCTION

Organophosphorus molecules, including warfare agents and pesticides, are highly toxic compounds that can pose serious risks to human health and the environment, including neurotoxicity and contamination of water resources, so their efficient removal is important for environmental priority, as well as for removal of warfare agents in defensive technologies. The study on the impact of graphene-based material structure shows that adsorption performance depends strongly on the nature of the graphene surface, which makes graphene-based materials promising candidates for removing these pollutants from wate [1].

Graphene oxide-based materials have attracted considerable attention for removing organophosphorus pesticides from water because their adsorption behavior depends strongly on surface structure, oxidation state, and the preservation of the sp^2 carbon network. In particular, oxidized graphene surfaces favor adsorption of more polar aliphatic compounds such as dimethoate, while graphene basal planes with a more ordered π-electron system are more suitable for aromatic pesticides such as chlorpyrifos.

Recent review work on graphene oxide (GO)-based materials for neurotoxic organophosphates also emphasizes that GO and GO composites are highly effective because their oxygen-containing functional groups, defects, and tunable surface chemistry enabl multiple interaction pathways, including hydrogen bonding, electrostatic attraction, π–π interactions, and complexation. These features make GO a versatile platform for both adsorption and remediation of organophosphate pollutants in aqueous systems [2].

Besides experimental studies, density functional theory (DFT) can also be used to investigate these systems at the molecular level. DFT helps explain adsorption energies, charge transfer, preferred binding sites, and the role of functional groups and defects, which makes it a useful complement to experiments when interpreting why some graphene-based materials remove organophosphates more efficiently than others [3].

The goal of the work was to establish a simple, general DFT-based methodology for investigation of adsorption of organophosphate molecules on graphene oxide, including warfare agents and pesticides. Tabun, as the representative of warfare agents, and dimethoate and malathion, as the representatives of pesticides, were selected for this investigation This is expected to provide insight into the general features and adsorption properties for structurally similar compounds.

2. CALCULATION DETAILS

Density functional theory (DFT) calculations were used to investigate the interaction between the organophosphate

Corresponding author: D. Vasić Anićijević
E-mail: dragana.vasic87@gmail.com

DOI: 10.5937/OTEH26118V

molecule and the graphene oxide (GO) surface at the molecular level. The computational approach was chosen to provide insight into the adsorption mechanism, the strength of interaction, and the role of oxygen-containing functional groups and surface defects in stabilizing the adsorbate on the GO sheet.

DFT calculations were performed using DFTB+ code [4] using 3ob-3-1basis set. The calculation procedure began with the construction of the isolated GO model, the isolated organophosphate molecule, and the combined adsorbate–adsorbent system. Graphene oxide was modelled as a 128-atom graphene slab with 8 oxygen atoms in epoxy positions and 8 OH-groups in C-top positions to simulate oxygen-rich environment. At least 30 angstroms of vacuum above the slab were provided to minimize interactions between periodic images. All structures were then geometry optimized to obtain the most stable configuration and to ensure that the total energy corresponded to a minimum on the potential energy surface. All atoms were moved through relaxation and 300 geometry optimization steps were allowed. During optimization, the electronic convergence criteria10^{-6} eV and force thresholds 5×10^{-3} eVA^{-1}) were applied to guarantee reliable structural and energetic results.

After geometry optimization, the adsorption energy was calculated according to the following expression:

$$E_{ads}=E_{complex}-E_{GO}-E_{OPs} \quad (1)$$

where $E_{complex}$ is the total energy of the GO–organophosphate system, E_{GO} is the energy of the isolated graphene oxide surface, and E_{OPs} is the energy of the isolated organophosphate molecule. A negative adsorption energy indicates a thermodynamically favorable interaction, while a more negative value corresponds to stronger binding between the molecule and the GO surface.

3. RESULTS AND DISCUSSION

Optimized adsorption geometries os DMT, MLT and GA on GO are represented in Figures 1-3. Multiple input molecular orientations were selected to favorize the potential interaction between heteroatom moieties of organophosphate molecules (P=S, N-H and O-H binding), as well as large surface contact areas (Van der Wals forces). Structurally, the GO lattice remains intact across the configurations, preserving its hexagonal carbon framework and the distribution of hydroxyl, epoxy, and carboxyl groups that facilitate adsorption. Across all tested organophosphorus compounds (dimethoate (DMT), malathion (MLT), and tabun (GA)), the adsorption energies (E_{ads}) reveal a consistent trend of strong binding to the graphene oxide surface. Each molecule exhibits negative adsorption energies, confirming thermodynamically favorable interactions.

Stable adsorption is attributed both to physical (hydrogen bonding, Van der Wals, electrostatic) and chemical interactions (S-O and N-H binding between substrate and adsorbate). Furthermore, all three compounds demonstrate multiple stable adsorption configurations with similar adsorption energetics, highlighting the structural flexibility of GO in accommodating different orientations and binding modes.

Despite these commonalities, notable differences emerge in adsorption strength and orientation (**Table 1**). Dimethoate (DMT) shows the most stable adsorption overall, with DMT-3 reaching the lowest energy (−294.55 kJ mol^{-1}), indicating strong surface coupling through both N-H-O and C-O-H interactions (hydrogen bonding). In case of MLT-2, S-O binding is the most obvious in adsorption geometry. Malathion (MLT) exhibits weaker adsorption in MLT-1 and MLT-2 (−217.43 and −213.62 kJ mol^{-1}), but achieves significantly stronger binding in MLT-3 (−264.12 kJ mol^{-1}), suggesting the significant contribution of hydrogen binding and Van Der Wals interactions to the overall binding energy. Similarly, Tabun (GA) presents the widest variation upon the change of adsorbate-to-substrate contact area: GA-1 shows the weakest adsorption (−186.97 kJ mol^{-1}), while GA-2 demonstrates a much stronger interaction (−266.18 kJ mol^{-1}), reflecting how positional differences (flat vs. upright orientation) drastically alter binding affinity.Collectively, these results underscore that adsorption energy is highly sensitive to molecular geometry, with DMT generally outperforming MLT and GA in stability, while GA exhibits the largest configurational variability.

Table 1. Adsorption energies (E_{ads}) of malathion, dimethoate and tabun on graphene oxide, for investigated adsorption geometries.

OPs		E_{ads} (kJ mol^{-1})
dimethoate	DMT-1	-285,39774
	DMT-2	-255,07476
	DMT-3	-294,55398
malathion	MLT-1	-217,43244
	MLT-2	-213,61734
	MLT-3	-264,11796
tabun	GA-1	-186,96816
	GA-2	-266,18094

Table 1 presents the adsorption energies (Eads) of three organophosphorus compounds: dimethoate (DMT), malathion (MLT), and tabun (GA) in different adsorption configurations on graphene oxide (GO). All values are negative, confirming that the adsorption processes are thermodynamically favorable.

Optimized adsorption geometries of DMT, MLT and GA on graphene oxide are represented in **Figures 1-3**.

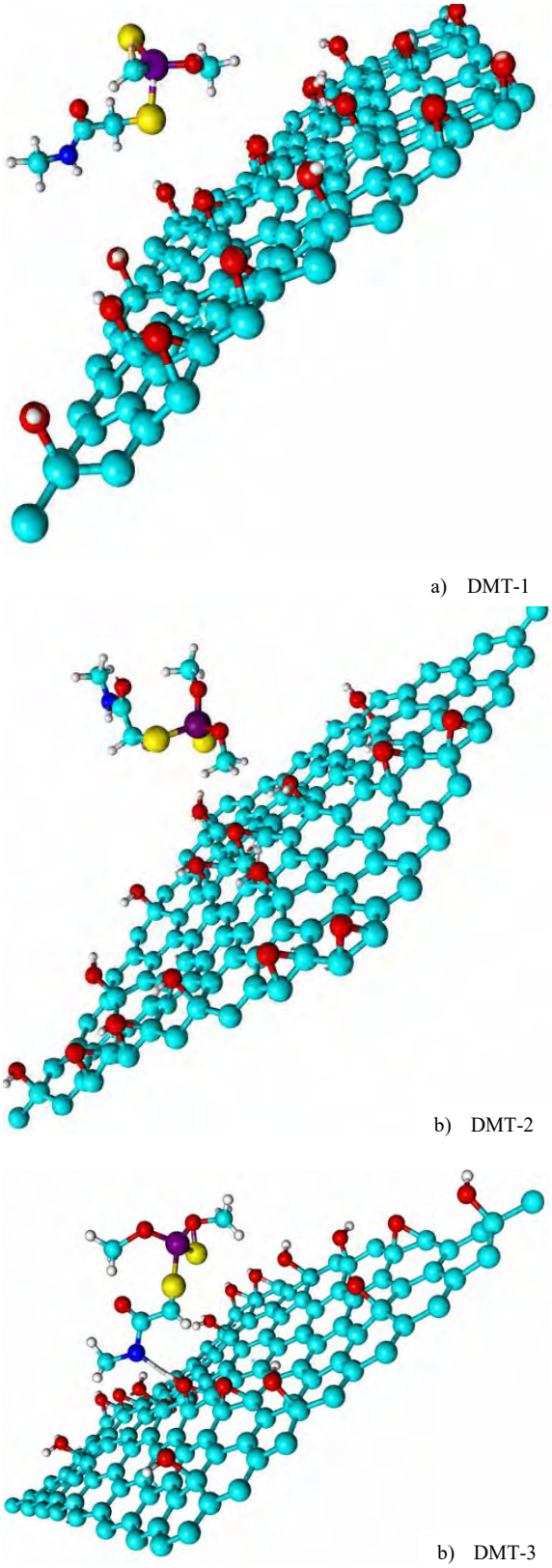


Figure 1. Optimized adsorption geometries of DMT on graphene oxide

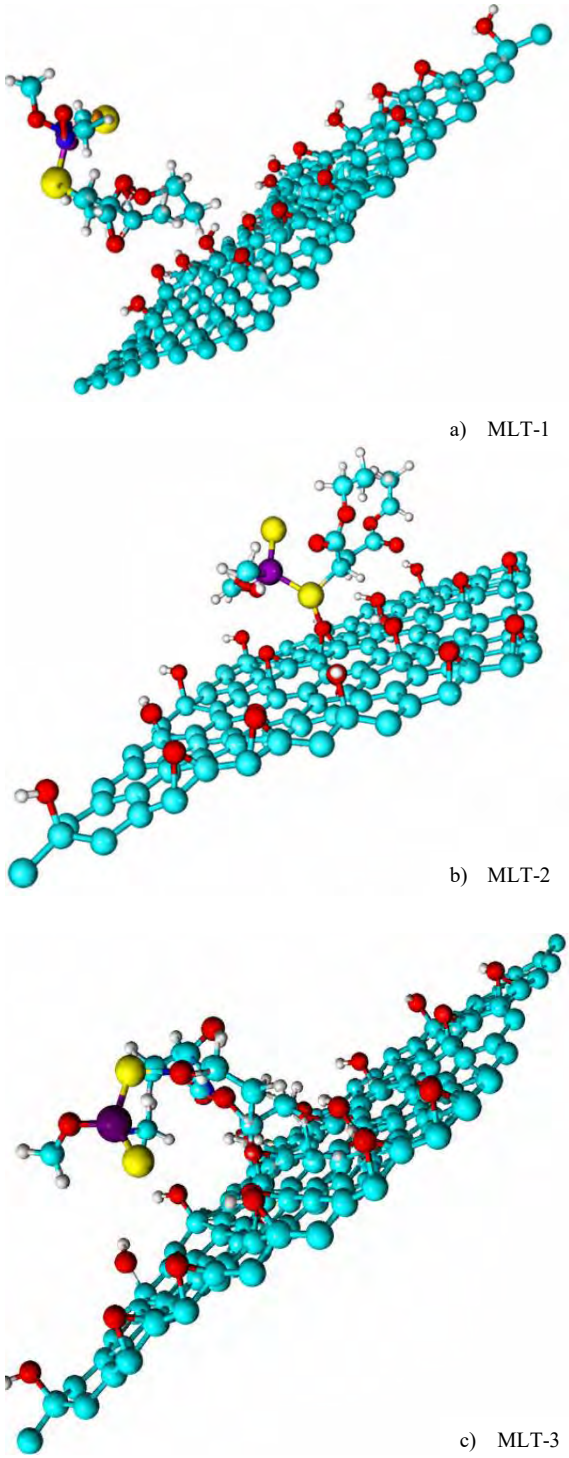
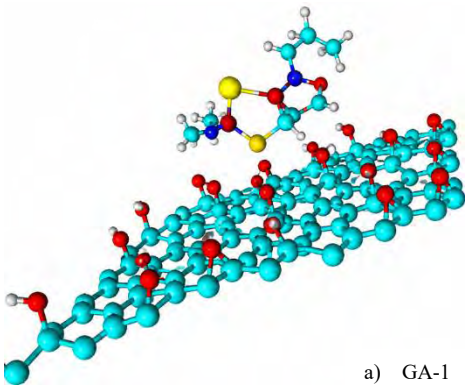
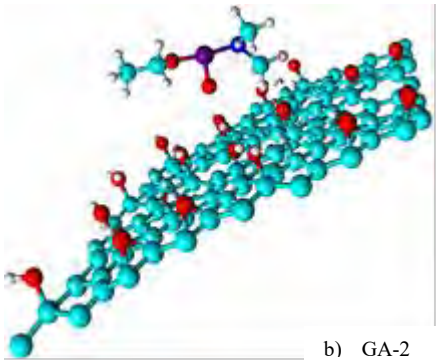


Figure 2. Optimized adsorption geometries of MLT on graphene oxide



a) GA-1



b) GA-2

Figure 3. Optimized adsorption geometries of GA on graphene oxide

4. CONCLUSION

The results presented in Table 1 and illustrated in Figures 1–3 clearly demonstrate that adsorption of organophosphorus compounds on graphene oxide is energetically favorable, with all calculated adsorption energies being negative. Among the studied systems, dimethoate (DMT) exhibits the strongest and most stable adsorption, particularly in configuration DMT-3, highlighting the importance of molecular orientation in maximizing surface interactions. Malathion (MLT) shows moderate adsorption energies, with a significant increase in stability in MLT-3, confirming that steric effects and binding geometry strongly influence adsorption strength. Tabun (GA) displays the widest variability, ranging from weak adsorption in GA-1 to markedly stronger binding in GA-2, underscoring the sensitivity of this compound to configurational changes.

Overall, the combined numerical and structural analyses indicate that graphene oxide provides a versatile platform for binding organophosphorus agents, with adsorption strength governed primarily by the spatial arrangement of phosphorus and sulfur centers relative to oxygen rich functional groups. These findings emphasize the potential of GO as an effective adsorbent material for environmental remediation and decontamination applications.

Acknowledgment

This work was partially supported by the Ministry for Science, Technological Development and Innovation of the Republic of Serbia (451-03-33/2026-03/200017 and 451-03-34/2026-03/200146 and Ministry of defence Republic of Serbia (VA-TT/1/22-24).

References

[1] T. Lazarević-Pašti, V. Anićijević, M. Baljozović, D.V. Anićijević, S. Gutić, V. Vasić, N.V. Skorodumova, I.A. Pašti, The impact of the structure of graphene-based materials on the removal of organophosphorus pesticides from water, Environmental Science: Nano 5(6) (2018) 1482-1494.

[2] V. Anićijević, T. Mitrović, T. Terzić, T. Lazarević-Pašti, Graphene Oxide-Based Materials for the Remediation of Neurotoxic Organophosphates, Processes 13(12) (2025) 4028.

[3] A. Ishii, K. Nakada, DFT Calculation for Adatom Adsorption on Graphene, in: J.R. Gong (Ed.), Graphene Simulation, IntechOpen, London, 2011.

[4] B. Hourahine, M. Berdakin, J.A. Bich, F.P. Bonafé, C. Camacho, Q. Cui, M.Y. Deshaye, G.D. Mirón, S. Ehlert, M. Elstner, Recent developments in dftb+, a software package for efficient atomistic quantum mechanical simulations, The Journal of Physical chemistry. A 129(24) (2025) 5373.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ADDITIVE MANUFACTURING IN THE MILITARY INDUSTRY: RESILIENCE AND SUSTAINABILITY FOR SMALL STATES

SAŠA VESELINOVIĆ,
Technical Test Center of Serbian Armed Forces, 445 Vojvode Stepe, 11000 Belgrade, sasa.veselinovic@tutanota.com

STEVAN JOVIČIĆ,
Technical Test Center of Serbian Armed Forces, 445 Vojvode Stepe, 11000 Belgrade, stevanjovicic@gmail.com

DRAGAN STOJADINOVIĆ,
Technical Test Center of Serbian Armed Forces, 445 Vojvode Stepe, 11000 Belgrade,
dragan.stojadinovic90@gmail.com

Abstract: *Military industries operate with many opportunities, but also face obstacles in the form of political, economic or security restrictions. This paper analyzes additive manufacturing, or 3D printing, as a technology which can provide more resilient production, development, and sustainability of military capability for small states. By analyzing open-source strategic reports and scientific studies, the paper identifies three interlinked dimensions in which 3D printing can contribute to sustainable military production under various political and security conditions: security resilience, economic efficiency and environmental responsibility. The findings show that additive manufacturing can reduce dependence on imports of different components, which can mitigate challenges related to international sanctions, export restrictions or supply chain disruptions. Analyzed sources indicate that 3D printing can offer greater adaptability, lower cost, and a faster development process for small scale production. It can reduce material waste, energy consumption, transportation costs and carbon emissions. The global trends, opportunities and present risks, such as patent rights, cybersecurity and arms proliferation control are taken into consideration. Policy recommendation includes establishment of 3D printing hubs, product modification strategies and the adoption of energy efficiency and ecology standards. Overall, 3D printing can provide a higher degree of autonomy and resilience in military production for small states.*

Keywords*: additive manufacturing, 3D printing, military industry, production sustainability, small states, production efficiency.*

1. INTRODUCTION

Nowadays, military industries are frequently hit by political disturbances, economic obstacles, security threats, and ecological catastrophes. These constraints are usually expressed in forms of political requests, international sanctions, weapon embargoes, or trade tariffs.

Production of military weapons and equipment in small states is usually based on massive imports of different components and spare parts. The supply is relied on long-distance supply chains, centralized supply routes, suppliers from various countries, and a long time for execution. These characteristics make the supply very sensitive to any possible constraints.

The military industry can use developed technology of additive printing (AM) or 3D printing, as a national defense strategy policy to establish operational resilience under disruptive conditions [1]. This technology offers opportunities for locally produced components, in small series with small-scale material stocks, using universal 3D printers and DIGITAL FILE models. It constructs objects by adding material layer by layer, with minimal leftover. Modern technology produces very complex components technology can produce very complicated components, consisting of different materials, depending on printing technology [2].

According to open-source strategic reports and case studies, AM contributes to the resilience of military production in small states through three areas: security resilience, economic efficiency, and environmental responsibility.

Along the positive sides of AM technology, the paper discusses critical risks, focusing on the patent rights, cybersecurity, arms proliferation control, and required precondition for locally 3D printing of spare parts [3].

In conclusion, the paper translates findings into several policy recommendations, including the establishment of regional 3D printing hubs, on-site modification workshops, and the implementation of material recycling standards in manufacturing.

Corresponding author: S. Veselinović;
E-mail: sasa.veselinovic@tutanota.com
DOI: 10.5937/OTEH26119V

2. BASIC CONCEPTS: ADDITIVE MANUFACTURING, MILITARY INDUSTRY, AND INTERNATIONAL TRADE CONTROLS

Characteristics of AM, the production of weapon systems, and international trade conditions are in correlation to component supplies. Each domain introduces analytical categories that are essential for interpreting empirical evidence presented in subsequent sections of this paper.

2.1. Additive Manufacturing: Definition, Processes, and Characteristics

Additive manufacturing (AM) is an innovative technology where the object creation is based on layer-by-layer material addition, in contrast to conventional manufacturing, where the object creation is based on removing excess material.[2].

For small series component production, AM offers several advantages over conventional manufacturing: almost no material waste, easy fabrication, minimal human work, little post-processing, and energy efficiency. [4]

These properties make 3D printing relevant for production in the military industry, for small-scale production, where availability of components is crucial, the time of supply is essential, and the cost reduction is fundamental. The Peace Research Institute Frankfurt characterizes AM as "key for the development of future military capabilities" and notes that the technique "produces objects layer by layer according to a digital 3D model" [5].

According to the American Society for Testing and Materials (ASTM) classification, AM processes are divided into seven main categories [4]: Stereolithographic (SL), the earliest process, introduced and patented by Charles Hull in 1984; Fused Deposition Modelling (FDM), developed by Stratasys; Powder Bed Fusion (PBF), encompassing Selective Laser Sintering (SLS), Electron Beam Melting (EBM), and Selective Laser Melting (SLM); Binder Jetting, developed at MIT; Direct Energy Deposition (DED), specifically employed for repair and maintenance rather than new-part fabrication; and Laminated Object Manufacturing (LOM), commercialized by Helisys Inc. in 1991. Every AM technology has different characteristics and fields of application.

2.2. Technological Framework

Additive manufacturing is strongly connected with industrial development aligned with Industry 4.0. Raja Santhi, A. and Muthuswamy, P. [6] describe Industry 4.0 as an industry transformation based on digital connection and data-driven processes, focused on automation, interconnectivity, and real-time process optimization. In this environment, leading technologies are the Internet of Things (IoT), Machine Learning (ML), Artificial Intelligence (AI), Cyber-Physical Systems (CPS), Cloud Computing, Additive Manufacturing, Digital Twins, and Cybersecurity. All of these technologies from Industry 4.0 create a manufacturing ecosystem that is interconnected and data-driven [6].

On the other hand, military industry production has always been expensive, involved high technology, complicated product development, and secrecy. In the era of globalization, the costs of development and production have been reduced by outsourcing some work or by buying some components from the main world producer. Such a solution leads to many challenges: trade barriers, disruption of supply-chains, security challenges, time for delivery, quality control, etc. In some cases, for high-tech systems, for small-scale production, the low price is not the main goal, but the availability of components on time.

2.3. International Trade Controls, Sanctions, and Trade Barriers

Hirbu [7] provides a foundational definition: "A trade restriction is defined as any government policy that limits the free flow of goods and services across borders. The main goal of trade restrictions is to protect domestic industries from cheap imports from other countries".

Of greater relevance to the military-technological context is the author's typology of four categories of formal trade barriers: Voluntary Export Restraints, Regulatory Barriers, Anti-Dumping Duties, and Subsidies [7]. Beyond formal tariff barriers, the literature identifies non-tariff measures as increasingly central instruments of trade governance. Hirbu [7] notes that "non-tarife barriers are way to restrict trade using trade barriers in a form other than a tariff ... There are different types of non-tariff measures, such as quotas, sanctions, import deposits, licenses, embargoes, etc."

Monadjemi and Lodewijks [8] highlight the strategic, power-political dimension of trade restrictions: "There is a literature on strategic trade policy, and this appears a power play to reduce market access to improve a bilateral trade deficit." This insight is relevant for understanding how sanctions and export controls function, not merely as regulatory instruments but as tools of geopolitical competition, directly affecting the availability of advanced manufacturing technologies, including AM systems, to states subject to such measures.

3. METHODOLOGY AND ANALYSIS OF THE THREE CORE DIMENSIONS

3.1. Security Resilience

The primary utility of AM within a small state's defense strategy is its capacity to transform a physical, vulnerable logistics pipeline into an anti-fragile, digital asset protection network.

Traditional weapons embargoes and economic sanctions rely on physical interdiction at critical geographical choke points such as maritime ports, airspace corridors, and border crossings. Additive manufacturing fundamentally alters this interdiction dynamic by digitizing the transaction layer, and they become invisible to traditional

maritime blockades and customs border checks [9].

Field data from international deployments confirms that the speed of local component execution acts as a major resilience factor during active disruptions. In expeditionary or isolated defensive settings, standard logistical transport pipelines typically feature severe vulnerabilities due to damaged local infrastructure or a lack of secured transport [2].

3.2. Economic Efficiency

Military supply chains during operations differ from civilian supply chains in ways that amplify the vulnerabilities of smaller states. Den Boer et al. [2] identify three distinguishing features: the primary focus is not on financial efficiency but on advancing the operational effectiveness of deployed troops; the circumstances in which troops operate are of considerably greater variety; and during military and humanitarian operations, the supply chain frequently encounters damaged infrastructure, lack of transport, loss of governmental functions, threat or physical violence, and the presence of many injured and traumatized individuals.

Westerweel et al. [16] conclude that on-site 3D printing can help to solve some of the biggest issues. It is presented the new model formulation in which a 3D printer can be used to resolve spare parts shortages. The benefits of on-site printing are demonstrated via a case study based on actual Royal Netherlands Army operations in Mali. The results prove that RNLA greatly benefits from on-site printing, much more than from a conventional expediting option. It shows a reduction in operational costs of 47% across 14 components, and inventory storage space for regular spare parts is reduced by 72%. Those benefits are achieved with a relatively simple general-purpose printer against small investments in on-site raw material [16].

3.3. Resource saving and recycling

The military industry is structurally optimized for weapon production, where the raw material is optimized for subtractive production, ignoring ecological costs of raw material extraction, scrap generation, and intercontinental transportation loops [10]. In conventional subtractive production, up to 80% to 90% of the raw material is ground away into unrecyclable shavings, which becomes a massive waste footprint for premium, energy-intensive materials [1].

Additive manufacturing has different material processing based on different technology [12]. 3D printers build components layer-by-layer, using the precise volume of printing powder required for the component's geometry [13]. Any unspent material powder surrounding the build chamber is vacuum-reclaimed, sieved, and recycled into subsequent production runs, capping net material waste below 5% to 10% [13]. For a small state facing import restrictions, this 90% reduction in material scrap ensures that precious stocks of national raw metals are preserved rather than lost to factory floor waste.

Official field deployment records for the Dutch troops in the UN mission in Mali indicate that over 51% of all freight cargo delivered to remote military bases consists of single-use packaging materials, polymer containers, and protective sealing foils, with less than 0.1% of these materials ever returned to the home country [2].

By deploying ruggedized shredders and compact filament extruders within the base infrastructure, single-use polymer packaging can be processed locally into technical-grade printing filaments [2]. Processing the baseline 51% packaging waste stream available on-site drops external supply dependency down to 56.65%, proving that on-site circular recycling acts as a strategic buffer [2]. This transformation turns military production into a localized, self-sustaining loop, significantly reducing the carbon footprint of transport vehicles while securing parts supply without drawing on external resources [11].

4. GLOBAL TRENDS, RISKS, AND CHALLENGES

4.1. Resilience policy trends in study cases

The case of the Royal Netherlands Army is particularly instructive as an example of how a small but technically advanced NATO member has approached the AM-logistics nexus. Den Boer et al. document that "the Dutch Army has been going through several budget cuts in the last decade. It has caused the military to be feeding off reserves, both financially and physically in spare parts and material. These budget cuts have led to situations in which availability requirements of materials and spare parts for certain units were not met, in turn affecting preparatory activities for missions and on combat power during missions themselves" [2].

The institutional response was equally significant: "As part of the government strategic plan, the Dutch Ministry of Defense is supposed to reduce primary raw material use by 50%" [2]. This commitment simultaneously addresses economic, environmental, and operational objectives, with AM as the enabling technology.

The most extreme contemporary illustration of AM as a supply chain resilience instrument under conditions of acute disruption is provided by the Ukrainian wartime context. Suckau describes how "both industrial actors and private citizens form nodes in a crowdsourced supply network that supplies the army. As such, it not only shows how lethal effects are achieved, compared to industrial applications' simpler 3D printing, but also exemplifies the professionalization and adaptation of an ad hoc innovation that initially responded to a lack of materials and supply. Besides polymer-based products, replacement parts for conventional armaments have been printed with machines provided through international military assistance".[5] This case transcends the conventional state-centric model of military logistics, demonstrating that under conditions of extreme supply chain disruption, AM enables decentralized, distributed production networks involving civil society actors.

The author appropriately notes the epistemological

limitation of this case: "It is too early to assess the impact on Ukrainian military capabilities" [5]. Nevertheless, the case provides a documented illustration of the mechanisms through which AM can substitute for or supplement conventional supply chains when those chains are partially or fully disrupted.

4.2. Intellectual Property (IP) and Legal Vulnerabilities

When there are supply chain disruptions, regardless of whether they are political isolation, embargoes, or blockades, it immediately affects military production and defense forces readiness [2]. To keep weapon production and maintenance alive by using reverse engineering and locally 3D printing components without authorization from foreign original equipment manufacturers. This can be an immediate solution for defense capabilities, but long-term it can induce problems with contractual and guarantee breachs, but also can cause additional diplomatic countermeasures.

As documented by defense logistics reviews, utilizing self-produced spare parts instead of OEM-certified components routinely invalidates overall system warranties. This can lead to intense legal disputes regarding liability if a system fails during operations [2].

The international community and companies will not look favorably on the infringement of patent rights, such as the mass production of parts without a license. This will have as a consequence secondary economic sanctions, further trade restrictions, or military actions in some cases, which can be extremely harmful for the military structures of small states.

4.3. Cyber-Physical Security Vulnerabilities

AM infrastructure is relying on digital design files, which opens up for hostile cyber operations [9], especially when the CAD software and 3D printers are usually produced by global foreign companies. In some cases, 3D printers are connected to the Internet to their producer, from whom the printing file has to be received. Adversaries can use these facts to cyber-physically sabotage.

Cyber attackers do not need to destroy a network or prevent printing components. Instead of that, the printing file can be intercepted, and the component registries can be changed. These changes can lead to small changes in design or printing quality, which can have an impact on printed component function [3].

These compromised parts will visually pass field inventory checks, only to suffer catastrophic failure under operational stress (e.g., a drone propeller snapping at high RPM or a missile casing rupturing upon ignition), rendering the weapon system ineffective at a critical moment.

4.4. Technical Omissions and Quality Control

Maintaining consistent mechanical properties across distributed, forward-deployed 3D printers remains a difficult engineering challenge. In centralized mass production, quality control is maintained through tightly regulated factory conditions [1]. On the field, however, AM systems must operate in harsh, volatile environments. Institutional reports from the U.S. GAO and the NATO STO [14] mention uncontrolled changes in humidity, and fluctuations in electrical grids, but also the quality of trained personnel.

Uncontrolled changes in humidity, ambient temperature, and dust contamination can degrade raw metal powders and technical polymers [1]. For example, nylon and advanced polyamides are highly sensitive to moisture, and poor storage conditions can lead to porous, brittle prints [2].

Fluctuations in electrical grids or rough physical handling of machines can throw off optical laser calibration or extrusion precision [1]. This can affect the precision of the production head, which can be a big problem for high-precision military systems [2].

The problem of trained personnel who are responsible for printing components or quality control is problematic because usually it’s their additional job, and the different components have different requirements. It’s difficult to train personnel for every possible situation.

4.5. Arms Control and Proliferation Challenges

Today, 3D printing is a widely spread technology, with very capable 3D printers available for a relatively low price. If CAD printing file reach a person with criminal intentions, they can easily produce components of highly capable weapon systems. Traditional arm control systems are based on tracking physical components, and they aren’t successful. Tracking digital files is much harder, and it’s dangerous.

UNIDR [3] highlights that AM deconstructs this framework by separating the design of a weapon from its material execution. Unindexed weapon designs, modular CAD frameworks for improvised explosive devices (IEDs), and components for autonomous kamikaze drones can be shared globally across untraceable digital networks.

This digital transfer makes it increasingly difficult for international bodies to enforce traditional export controls or verify non-proliferation treaties. This environment allows both non-state actors and domestic criminal enterprises to establish covert, localized manufacturing cells that bypass all conventional border and customs checkpoints [9].

5. POLICY RECOMMENDATIONS

It isn’t possible to set one solution for building resilience of military industry production in small states, considering global trade market disturbances. To get the most from AM and mitigate the system risks, the best solution for small states is to establish national policies, with the aim to build military industry, including AM as integrated technology. That will include the establishment of regional 3D printing hubs, product development aiming to design for 3D printing and required ownership of data

files, production standardization, and required procedures for material recycling, storage, and testing.

5.1. Establishment of Sovereign and Secure 3D Printing Hubs (AM Hubs)

Small states should avoid fragmented, ad-hoc adoption of AM by civilian operators and instead build a centralized network of secure, military-controlled Additive Manufacturing Hubs (AM Hubs) [3]. These installations must serve as the primary repositories for national digital inventory data, technical materials, and industrial-grade printing systems (e.g., Wire Arc Additive Manufacturing - WAAM and Selective Laser Melting - SLM) [12].

AM Hubs should be built as production facilities with big data centers and different printing material reserves. It should operate as a service center for small-scale production where other types of production would demand too many operations or be too costly. As a production line, there should be many different 3D printers for printing all kinds of materials. For military purposes, this kind of facility can be built in cargo containers, and it could be set for expeditionary purposes.

As a data center for CAD file storage, this hub has to be protected from different cyber-physical vulnerabilities, so it has to be built with high network protection from physical impact and cyberattacks.

To insulate these hubs from the cyber-physical vulnerabilities detailed in Chapter 4, they must adhere to rigid network isolation criteria.

First, all computer-aided design (or other types of digital files) stations, slicing engines, and industrial printers must be completely disconnected from public internet routing and unclassified defense networks.

Data transfer from the main digital registry to forward-deployed or mobile printing modules must utilize high-level, end-to-end encryption protocols with hardware-token validation to prevent malicious code modifications or file interception.

Hubs should establish cooperation with national technology research centers and technical universities to provide an environment for technical improvement, material development, and 3D printing localization, but also to ensure a continuous pool of certified operators and design engineers without relying on foreign technical support.

5.2. Adoption of "Design-for-AM" and Digital Procurement Protocols

3D printing requires digital files, which are usually intellectual property and should be solved before the need for printing arises. Small states must pivot toward data-centric purchasing frameworks [2].

Future procurement contracts signed by small states with foreign original equipment manufacturers (OEMs) should mandate the following provisions: contractual technical data packages (TDPs), "Design-for-AM" optimization, and certified third-party service escrows.

Tenders must require that international suppliers include fully documented CAD files, material performance criteria, and localized printing rights for non-structural, high-wear components as part of the base purchase agreement.

Domestic defense industries must design native military equipment to prioritize the use of modular, standardized components that are explicitly optimized for on-demand 3D printing, reducing the need for specialized external tooling [1].

To protect against situations where an OEM goes bankrupt or imposes a political cutoff, agreements should place critical design blueprints in secure digital escrows managed by neutral international bodies or verified domestic service providers.

5.3. Codification of Digital Quality and Ecological Standards

For decentralized designing and manufacturing components

that will meet required structural performance and quality, but also the resource efficiency and ecological standards, small states must enforce strict, standardized quality controls [15].

Any critical component printed at a forward-deployed location must undergo strict verification protocols (e.g., ultrasonic scanning or industrial computed tomography) to check for internal microstructural voids and layer alignment anomalies before field installation [1].

Mobile, containerized printing labs must be equipped with active environmental control systems to manage internal moisture, temperature, and dust levels. This ensures raw metal powders and technical polyamides do not degrade before printing [1].

Small states should mandate for hubs to use high-efficiency machinery and compulsory recycling of unspent metal powders and single-use polymer packaging within tactical zones [11]. Standardizing on-site filament extrusion processes allows small states to transform base waste streams into functional defense assets, ensuring long-term resource independence under blockade [2].

6. CONCLUSION

Additive manufacturing is a disruptive technology, which can cause a structural shift in how small states design and maintain their defense capabilities and sustain national interests under geopolitical constraints. As global supply chain disruptions, caused by international sanctions, wars, export controls, or resource constraints, threaten the continuity of military industry production and degrade the defense capabilities of small countries. Conventional centralized supply models, characterized by long lead times and high inventory dependencies, increasingly fail to guarantee operational readiness.

This paper has evaluated the impact of 3D printing on small states military industry and defense capability.

Changing from physical import to transfer of encrypted digital data protects small states from physical supply chain breaks and embargoes, effectively bypassing traditional geographical choke points. AM excludes upfront tooling costs in conventional manufacturing and enables the financially viable production of rare, customized, and small-batch components, reducing weapon system downtime by over 90%. Building components layer-by-layer cuts net raw material waste below 10%, while localized circular loops (such as converting single-use base packaging waste into raw printing filament) allow outposts to sustain manufacturing needs without drawing new external resources.

Building a military industry around AM technology, it has its negative sides. The digitalization of supply chains is an opportunity for hostile actors to change printing files to cause undetected structural weapon malfunction. Unauthorized component replications have intellectual property and warranty challenges. For security reasons, the digitalization of weapon designs can cause severe problems for non-proliferation hurdles for global arms control.

Successful implementation of AM depends on proactive public policies. By establishing AM hubs, requiring digital data rights in procurement contracts, establishing strict control standards, and recycling and ecology standards, small states can mitigate these risks. Then, AM becomes an essential, sustainable ecosystem that provides small states with the operational flexibility, economic sustainability, and long-term autonomy required to survive in a volatile global arena.

References

[1] GOVERNMENT ACCOUNTABILITY OFFICE: *3D Printing: Opportunities, challenges, and policy implications of additive manufacturing (GAO-15-505SP)*, *Report to the Chairman, Committee on Science, Space, and Technology*, House of Representatives, Washington, USA, 2015

[2] BOER, D., LAMBRECHTS, J., KRIKKE, H.: *Additive manufacturing in military and humanitarian missions: Advantages and challenges in the spare parts supply chain*. Journal of Cleaner Production, 257 (2020), Article 120301

[3] GRAND-CLEMENT, S., REINERTSEN, S.S: *Additive manufacturing of conventional military equipment: Implications for arms control and security*, UNIDIR, Geneva, 2026.

[4] JANDYAL, A., CHATURVEDI, I., WAZIR, I., RAINA, A., HAQ M.I.U.: *3D printing - A review of processes, materials, and applications in industry 4.0*, Sustainable Operations and Computers, 3 (2022) 33-42.

[5] SUCKAU, L.: Additive manufacturing in the military technology sector: Application and proliferation of an emerged technology, Peace Research Institute Frankfurt, Report No. 3 (2025)

[6] RAJA SANTHI, A., MUTHUSWAMY, P.: *Industry 5.0 or industry 4.0S? Introduction to Industry 4.0 and a peek into the prospective Industry 5.0 technologies*, International Journal on Interactive Design and Manufacturing (IJIDeM), 17(2) (2023) 947-979.

[7] HIRBU, A.: *The role of trade restrictions and trade barriers in the contemporary world economy*. In Proceedings of the 17th International Scientific Symposium of Young Researchers, Academy of Economic Studies of Moldova (2019) (pp. 156-158).

[8] MONADJEMI, M., LODEWIJKS, J.: *Trade restrictions and economic development: A study of the effectiveness of trade barriers*. Expert Journal of Economics, *8*(1), (2020) 18-23.

[9] BROCKMANN, K., SIBYLLE, B.: *3D printing and missile technology controls: Dealing with digital technology transfers*, Stockholm International Peace Research Institute, SIPRI Research Report No. 47, 2017.

[10] KHAN, M., HALEEM, A., JAVAID, M.: *Changes and improvements in Industry 5.0: A strategic approach to overcome the challenges of Industry 4.0*, Green Technologies and Sustainability, 1(2) (2023) Article 100020.

[11] ATIF, S.: *Analyzing the alignment between circular economy and industry 4.0 nexus with industry 5.0 era: An integrative systematic literature review*. Sustainable Development, 31(4), (2023) 2155-2175.

[12] RAHMANI, R., KARIMI, J., RESENDE, P. R., ABRANTES, J. C. C., LOPES, S. I.: *Overview of selective laser melting for Industry 5.0: Toward customizable, sustainable, and human-centric technologies*, Machines, 11(5) (2023) 522.

[13] FORD, S., DESPEISSE, M.: *Additive manufacturing and sustainability: an exploratory study of the advantages and challenges*, Journal of Cleaner Production, 137, (2016) 1573–1587.

[14] NATO Science & Technology Organization: *Science & Technology Trends 2025-2045: Volume 1*. NATO STO Flagship Reports, 2025

[15] ROJEK, I.: *Energy efficiency and sustainability of additive manufacturing as a mass-personalized production mode in Industry 5.0/6.0*, Energies, 18(13) (2025) Article 3413.

[16] WESTERWEEL, B., BASTEN, R., DEN BOER, J., VAN HOUTUM, G. J.: Printing Spare Parts at Remote Locations: Fulfilling the Promise of Additive Manufacturing, Production and Operations Management, 30 (2020) 10.1111/poms.13298.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INTEGRATED ELECTROCHEMICAL ASSESSMENT OF STEEL CORROSION IN CHLORIDE ENVIRONMENTS

MILICA VLAHOVIĆ
University of Belgrade, Institute of Chemistry, Technology and Metallurgy-National Institute of the Republic of Serbia, Belgrade, m.vlahovic@ihtm.bg.ac.rs, mvlahovic@tmf.bg.ac.rs

MARINA DOJČINOVIĆ
University of Belgrade, Faculty of Technology and Metallurgy, Belgrade, rina@tmf.bg.ac.rs

ĐORĐE GJUMIŠEV
University of Belgrade, Institute of Chemistry, Technology and Metallurgy-National Institute of the Republic of Serbia, Belgrade, djordje.gjumisev@ihtm.bg.ac.rs, mvlahovic@tmf.bg.ac.rs

NATAŠA ĐORĐEVIĆ
Institute of General and Physical Chemistry, Belgrade, Serbia; natasa.djordjevic@iofh.bg.ac.rs

JOVANKA PEJIĆ
University of Belgrade, Institute of Chemistry, Technology and Metallurgy-National Institute of the Republic of Serbia, Belgrade, jovanka.kovacina@ihtm.bg.ac.rs

Abstract: *Electrochemical methods are widely used for evaluating corrosion behavior of structural steels in aggressive environments; however, their mutual consistency and practical applicability under varying conditions remain critical challenges. In this study, the corrosion behavior of X12CrMoWVNbN10-1-1 steel was investigated in 0.1 M NaCl solutions at different pH values (2.9, 6.5, and 10.3) using electrochemical impedance spectroscopy (EIS), linear sweep voltammetry (LSV), and potentiodynamic polarization (PDP), complemented by scanning electron microscopy (SEM). The results reveal a pronounced influence of pH on both corrosion rate and degradation mechanism. Acidic conditions promote active dissolution with high corrosion current density, while neutral and alkaline environments enable passive film formation, resulting in significantly improved corrosion resistance. A combined analysis of electrochemical techniques indicates good agreement between EIS and LSV in less aggressive conditions, whereas notable deviations were observed in acidic media due to accelerated corrosion kinetics and interfacial instability. PDP measurements further confirm enhanced resistance to pitting corrosion at higher pH values. Surface analysis supports the electrochemical findings, showing severe degradation in acidic solution and compact, stable surfaces in neutral and alkaline environments. The study demonstrates that the combined application of electrochemical and surface characterization methods provides a reliable framework for corrosion assessment and highlights the importance of method selection in engineering practice.*

Keywords: *electrochemical impedance spectroscopy; linear sweep voltammetry; potentiodynamic polarization; pH effect.*

1. INTRODUCTION

Corrosion of structural steels operating in chloride-containing environments remains one of the major challenges in engineering practice because it directly affects the reliability, service life, and maintenance costs of critical components [1,2]. Chloride ions are particularly aggressive since they promote passive film breakdown and initiate localized corrosion, while environmental pH strongly influences both the corrosion rate and the stability of protective oxide layers [3,4].

Electrochemical techniques are among the most effective tools for evaluating corrosion behavior because they provide rapid, quantitative, and non-destructive information on metal degradation [5,6]. Electrochemical impedance spectroscopy (EIS), linear sweep voltammetry (LSV), and potentiodynamic polarization (PDP) are widely employed for determining corrosion resistance, corrosion kinetics, and susceptibility to localized corrosion under various environmental conditions [5,7-9]. However, each technique characterizes a different aspect of the electrochemical process, and their combined interpretation often provides a more reliable assessment than the application of a single method [6,8].

Ferritic–martensitic steels are extensively used in thermal power plants and other high-temperature engineering systems owing to their excellent mechanical properties and satisfactory corrosion resistance. Among them, X12CrMoWVNbN10-1-1 steel is widely applied for components operating under demanding service conditions [10]. Although its corrosion behavior has been investigated in different environments, studies comparing

Corresponding author: M. Vlahović;
E-mail: m.vlahovic@ihtm.bg.ac.rs;
DOI: 10.5937/OTEH26120V

the complementary application of electrochemical techniques for corrosion assessment under varying pH conditions remain limited [11,12].

The objective of the present study is therefore not only to evaluate the corrosion behavior of X12CrMoWVNbN10-1-1 steel in chloride solutions with different pH values, but also to compare the information obtained by EIS, LSV, and PDP measurements and correlate these findings with scanning electron microscopy observations. Particular attention is devoted to the complementary role of the applied techniques in assessing general and localized corrosion, with the aim of establishing an integrated electrochemical assessment strategy applicable to engineering practice.

2. EXPERIMENTAL

2.1 Material and Test Solutions

The investigated material was ferritic-martensitic steel X12CrMoWVNbN10-1-1 (EN 1.4906), commonly used for high-temperature components in thermal power plants because of its excellent mechanical properties and good corrosion resistance. Prior to electrochemical testing, the specimens were mechanically ground using SiC abrasive papers up to 1500 grit, degreased with ethanol, rinsed with distilled water, and dried in air.

Corrosion tests were performed in 0.1 M NaCl solution under three different environmental conditions: acidic (pH 2.9), neutral (pH 6.5), and alkaline (pH 10.3). The pH values were adjusted using diluted hydrochloric acid or sodium hydroxide solutions and verified before each experiment.

2.2 Electrochemical Measurements

Electrochemical measurements were carried out using a Gamry Reference 620 potentiostat/galvanostat in a conventional three-electrode electrochemical cell consisting of the steel specimen as the working electrode (1 cm² exposed area), a saturated calomel electrode (SCE) as the reference electrode, and a platinum mesh as the counter electrode. All experiments were conducted after stabilization of the open-circuit potential at room temperature.

Electrochemical impedance spectroscopy (EIS) measurements were performed by applying a sinusoidal perturbation of ±10 mV around the open-circuit potential over the frequency range from 100 kHz to 0.01 Hz. The impedance spectra were analyzed using an equivalent electrical circuit to determine the polarization resistance and parameters describing the electrochemical interface.

Linear sweep voltammetry (LSV) measurements were performed at a scan rate of 0.5 mV s^{-1} over the potential range of ±0.2 V with respect to the corrosion potential. The obtained polarization curves were used to determine the corrosion current density and corrosion rate by Tafel extrapolation.

The resistance of the investigated steel to localized corrosion was evaluated by potentiodynamic polarization (PDP). The measurements were carried out using the same electrochemical cell and scan rate, while the pitting potential and passive current density were determined from the anodic polarization curves.

All electrochemical measurements were repeated three times to verify the reproducibility of the obtained results.

2.3 Surface Characterization

After electrochemical testing, the morphology of the steel surface was examined using scanning electron microscopy (SEM). The observations were performed in order to correlate electrochemical parameters with the corresponding corrosion damage and to identify the characteristic features of passive film stability and localized corrosion under different pH conditions.

3. RESULTS AND DISCUSSION

3.1 Electrochemical assessment of general corrosion

The corrosion behaviour of X12CrMoWVNbN10-1-1 steel was evaluated by combining electrochemical impedance spectroscopy (EIS) and linear sweep voltammetry (LSV). Although both techniques are widely used for corrosion assessment, they provide complementary information on electrochemical processes occurring at the metal/electrolyte interface. EIS primarily characterizes the electrochemical response of the passive surface layer through impedance parameters, whereas LSV directly evaluates the kinetics of anodic dissolution and enables determination of corrosion current density.

EIS diagrams for X12CrMoWVNbN10-1-1 steel in 0.1 M NaCl solutions at different pH values are presented in Figure 1.

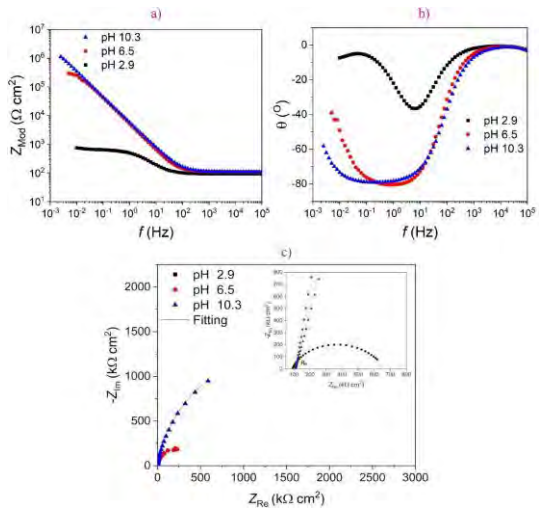


Figure 1. EIS diagrams for X12CrMoWVNbN10-1-1 steel in 0.1 M NaCl solutions at different pH values (pH 2.9, 6.5, and 10.3): (a) Bode modulus, (b) Bode phase, and (c) Nyquist plots

The impedance spectra obtained in chloride solutions (Figure 1) clearly demonstrate the strong influence of

solution pH on the corrosion resistance of the investigated steel. The acidic environment (pH 2.9) produced the lowest impedance modulus and polarization resistance, indicating rapid charge transfer and active metal dissolution. In contrast, both neutral and alkaline solutions exhibited considerably higher impedance values, reflecting the formation of a protective passive layer. The highest polarization resistance was obtained in the alkaline solution, confirming the beneficial effect of hydroxide ions on passive film stability.

The corrosion parameters obtained by LSV (Figure 2) followed the same overall trend of markedly higher corrosion resistance in neutral and alkaline solutions compared with the acidic environment. The highest corrosion current density was measured in the acidic solution, while corrosion rates decreased markedly in neutral and especially alkaline media. These results confirm that increasing pH significantly suppresses the general corrosion process by promoting passivation of the steel surface [4,12].

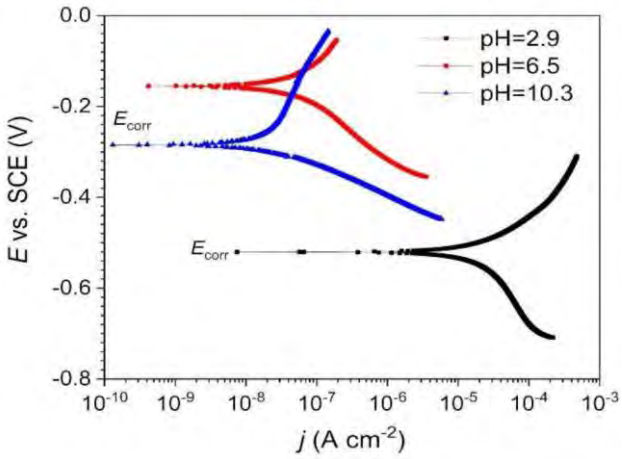


Figure 2. Tafel polarization curves for X12CrMoWVNbN10-1-1 steel obtained in acidic (pH 2.9), neutral (pH 6.5), and alkaline (pH 10.3) 0.1 M NaCl solutions

The polarization resistance values obtained from EIS and LSV are summarized in Table 1.

Table 1. Polarization resistance values obtained by EIS and LSV at different pH values.

Technique	R_p (kΩ cm^2)		
	pH 2.9	pH 6.5	pH 10.3
EIS	0.557	418.5	3051
LSV	2.391	507.7	1042

According to the presented results, both techniques yielded the lowest resistance in the acidic solution, confirming the substantially higher corrosion activity at pH 2.9. In neutral solution, the values were relatively close. In alkaline solution, however, a more pronounced difference was observed. Thus, the two methods agree in identifying the acidic environment as the most aggressive condition and the neutral and alkaline environments as considerably more protective, although they do not provide identical quantitative values or the same relative ranking of the neutral and alkaline conditions.

Such deviations originate from the different physical principles of the two methods [8,9]. While EIS evaluates the impedance response over a wide frequency range under near-equilibrium conditions, LSV determines corrosion parameters from polarization behavior during potential scanning. Consequently, LSV becomes more sensitive to rapid anodic dissolution occurring in highly aggressive acidic environments, whereas EIS reflects the overall electrochemical properties of the metal/electrolyte interface [5,6]. The comparison therefore indicates that the methods provide consistent information on the major differences in corrosion resistance, while their quantitative results should be interpreted as complementary rather than directly interchangeable.

The comparison of EIS and LSV results indicates that neither method alone provides a complete description of corrosion behaviour. Instead, their combined interpretation considerably improves the reliability of corrosion assessment, particularly under conditions where passive film stability strongly depends on environmental pH.

3.2 Evaluation of pitting susceptibility

General corrosion parameters alone are insufficient for evaluating the long-term performance of structural steels exposed to chloride-containing environments. For this reason, potentiodynamic polarization measurements were performed to assess the susceptibility of X12CrMoWVNbN10-1-1 steel to localized corrosion.

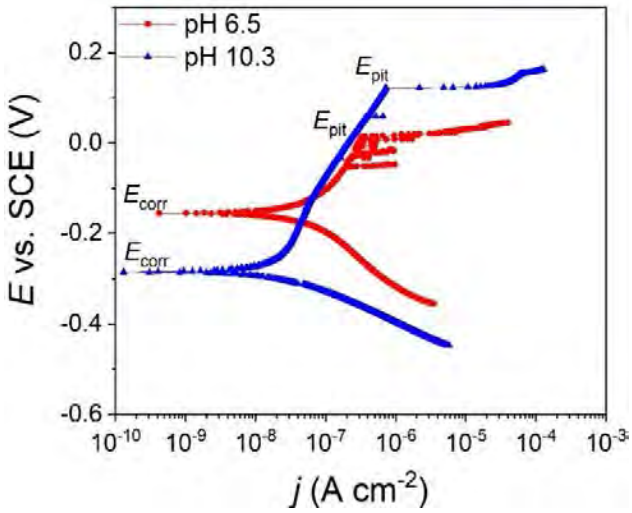


Figure 3. Potentiodynamic polarization (PDP) curves for X12CrMoWVNbN10-1-1 steel in neutral (pH 6.5) and alkaline (pH 10.3) 0.1 M NaCl solutions

The polarization curves (Figure 3) revealed a pronounced influence of pH on passive film stability and pit initiation [3,4]. In the acidic solution, active dissolution dominated throughout the anodic polarization range, indicating the absence of an effective protective film. Under these conditions, localized attack rapidly developed following continuous metal dissolution.

In neutral solution, the steel exhibited stable passivation over a broad potential interval before localized breakdown occurred. The passive region became even

more pronounced in the alkaline environment, where the highest pitting potential and the lowest passive current density were obtained. These findings demonstrate that alkaline conditions significantly increase the resistance of the investigated steel to chloride-induced pitting corrosion.

The PDP measurements complement the information obtained by EIS and LSV. Whereas EIS and LSV primarily quantify the intensity of general corrosion, PDP provides additional information regarding passive film stability and resistance to localized degradation. The combination of these electrochemical techniques therefore enables simultaneous evaluation of both uniform and pitting corrosion processes.

3.3 Correlation between electrochemical measurements and SEM observations

Electrochemical measurements provide quantitative information regarding corrosion kinetics and passive film behavior; however, direct observation of the exposed surfaces is essential for verifying the corresponding corrosion mechanisms. For this reason, SEM examination was performed after electrochemical testing.

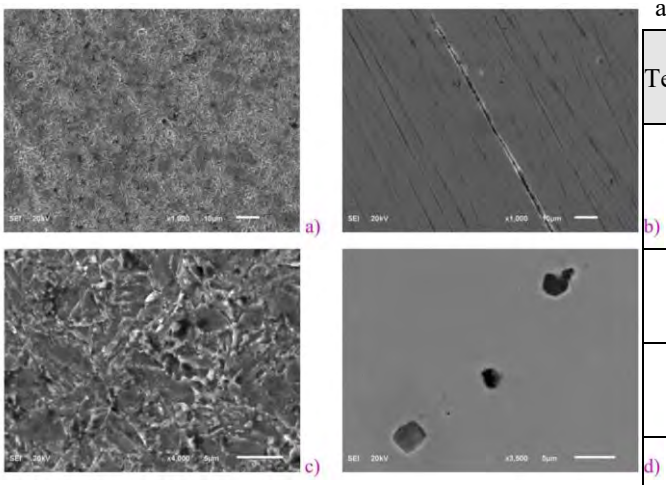


Figure 4. SEM micrographs of the X12CrMoWVNbN10-1-1 steel surface after exposure to 0.1 M NaCl: (a) pH 2.9 (×1000); (b) pH 6.5 (×1000); (c) pH 2.9 (×4000); (d) pH 6.5 (×3500)

The SEM micrographs (Figure 4) showed excellent agreement with the electrochemical results. Severe surface degradation was observed after exposure to the acidic chloride solution, where extensive dissolution and irregular corrosion features confirmed the high corrosion current density measured by LSV and the low polarization resistance obtained by EIS. These observations clearly indicate unstable electrochemical conditions associated with active corrosion.

Conversely, specimens exposed to neutral and alkaline solutions exhibited substantially less surface damage. In neutral solution, only isolated pits were detected, indicating localized passive film breakdown. The alkaline environment produced the most compact surface morphology, confirming the formation of a stable passive layer responsible for the highest polarization resistance and the lowest corrosion rate among the investigated conditions.

The close agreement between electrochemical parameters and SEM observations demonstrates that electrochemical measurements reliably describe the degradation processes occurring on the steel surface [13]. While EIS characterizes passive film resistance, LSV quantifies corrosion kinetics and PDP evaluates susceptibility to localized attack, SEM provides direct morphological confirmation of the electrochemical findings. Their combined application therefore represents a comprehensive methodology for corrosion assessment of structural steels operating in chloride-containing environments. Rather than identifying a superior technique, the present results demonstrate that reliable corrosion assessment is achieved through the complementary integration of EIS, LSV, PDP, and SEM, with each method contributing unique information on the corrosion process.

The complementary information provided by the applied corrosion assessment techniques is summarized in Table 2.

Table 2. Complementary information provided by the applied corrosion assessment techniques

Technique	Primary output	Information obtained	Role in corrosion assessment
EIS	Polarization resistance (R_p)	Passive film stability and corrosion resistance	Early evaluation of protective behavior
LSV	Corrosion current density (j_{corr})	General corrosion kinetics	Quantitative estimation of corrosion rate
PDP	Pitting potential (E_{pit})	Susceptibility to localized corrosion	Assessment of passive film breakdown
SEM	Surface morphology	Verification of corrosion mechanisms	Confirmation of electrochemical interpretation

4. CONCLUSIONS

The present study demonstrates that reliable corrosion assessment of X12CrMoWVNbN10-1-1 steel in chloride-containing environments requires the complementary application of several electrochemical techniques rather than reliance on a single measurement method. Although all investigated techniques consistently identified the beneficial effect of increasing pH on corrosion resistance, each method provided different but mutually complementary information on the corrosion process.

Electrochemical impedance spectroscopy proved particularly effective for evaluating passive film stability through polarization resistance measurements, while linear sweep voltammetry enabled quantitative determination of corrosion kinetics and corrosion rate. Potentiodynamic polarization provided additional information regarding the susceptibility of the investigated steel to localized corrosion by identifying

changes in pitting resistance under different environmental conditions. Scanning electron microscopy confirmed the electrochemical interpretation by revealing characteristic surface degradation features corresponding to the measured corrosion parameters.

From an engineering perspective, the obtained results demonstrate that acidic chloride environments promote active steel dissolution, whereas neutral and especially alkaline solutions significantly improve corrosion resistance through passive film stabilization. The excellent agreement between electrochemical measurements and surface observations confirms the reliability of the applied methodology.

Overall, the combined use of EIS, LSV, PDP, and SEM provides a comprehensive strategy for corrosion assessment of structural steels exposed to chloride environments and represents a practical approach for selecting appropriate evaluation methods in engineering applications.

Acknowledgements

The authors would like to thank the Ministry of Science, Technological Development and Innovations of the Republic of Serbia for support (Grant Nos: 451-03-33/2026-03/200026, 451-03-34/2026-03/200135, and 451-03-33/2026-03/200051).

References

[1] MUNDRA,S., SAMSON,G., MASI,G., ACHENBACH,R., BASTIDAS,D.,M., BERNAL,S.,A., BIGNOZZI,M.,C., CRIADO,M., CYR,M., GARTNER,N., VON GREVE-DIERFELD,S., LEGAT,A., NIKOONASAB,A., PROVIS,J.,L., RAUPACH,M., GLUTH,G.,J., G.: *Application of electrochemical methods for studying steel corrosion in alkali-activated materials*, Materials and Corrosion, 74(7) (2023) 988.

[2] HAARMANN,K., VAILLANT,J.,C., VANDENBERGHE,B., BENDICK,W., ARBAB,A.: *The T91/P91 Book*, 2nd ed., Vallourec and Mannesmann Tubes, USA, 1999.

[3] KATAYAMA,H.: *Evaluation of Corrosion Resistance of Coated Steels Using Electrochemical Methods*, Journal of the Japan Society of Colour Material, 88 (2015) 23.

[4] ASTM G102-89(2015)E1. *Standard Practice for Calculation of Corrosion Rates and Related Information from Electrochemical Measurements*, 2015.

[5] SEETHAMMARAJU,S., RANGARAJAN,M.: *Corrosion of stainless steels in acidic, neutral and alkaline saline media: Electrochemical and microscopic analysis*, IOP Conference Series: Materials Science and Engineering, Volume 577, International Conference on Advances in Materials and Manufacturing Applications, 577(1) (2018) 012188.

[6] RAMOS,C.,A., GONZÁLEZ,M.,B., BRUGNONI,L.,I., SAIDMAN,S.,B.: *Corrosion protection properties of polypyrrole coatings formed onto 316 L SS from a solution containing molybdate and salicylate*, Journal of the Serbian Chemical Society, 87(11) (2022) 1297.

[7] ZHANG,X.,M., CHEN,Z.,Y., LUO,H.,F., ZHOU,T., ZHAO,Y.,L., LING,Z.,C.: *Corrosion resistances of metallic materials in environments containing chloride ions: A review*, Transactions of Nonferrous Metals Society of China, 32(2) (2022) 377-410.

[8] SU,H., LUO,X., B., YANG,C.,F., CHAI,F., LI,H.: *Effects of Cu on Corrosion Resistance of Low Alloyed Steels in Acid Chloride Media*, Journal of Iron and Steel Research International, 21 (2014) 619-624.

[9] YUAN,J., CHEN,C., LI,J., QUAN, B., LAN,Y., WANG,L., FU,H., GAI,J.: *Initial Corrosion Behavior of 12Cr1MoV Steel in Thiosulfate-Containing Sodium Aluminate Solution*, Metals 10(10) (2020) 1283.

[10] FACHE,K., KEßLER,S., HARNISCH,J.: *On the Impact of Cathodic Polarization on the Chloride Threshold of Carbon Steel in Alkaline Solutions*, Materials and Corrosion, 76 (2025) 265–283.

[11] ASTM A335/A335M: *Standard Specification for Seamless Ferritic Alloy-Steel Pipe for High-Temperature Service*; ASTM International: West Conshohocken, PA, 2023.

[12] ELSENER,B., FANTAUZZI,M., ROSSI, A.: *Stainless steels: Passive film composition, pitting potentials, and critical chloride* content *in concrete*, Materials and Corrosion, 71(5) (2020) 797-807.

[13] PEJIĆ,J., KARPE,B., NAGODE,A., DOJČINOVIĆ,M., KOSEC,B., VLAHOVIĆ,M.: *Electrochemical and Surface Insights into X12CrMoWVNbN10-1-1 Steel Corrosion in Chloride Solutions at Variable pH*, Materials and Corrosion, 77 (4) (2026) 518-528.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

APPLICATION OF XRD AND AAS FOR EVALUATION OF MECHANOCHEMICAL REACTION PROGRESS

NATAŠA ĐORĐEVIĆ
Institute of General and Physical Chemistry, Belgrade, Serbia; natasa.djordjevic@iofh.bg.ac.rs

SRĐAN MATIJAŠEVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia; s.matijasevic@itnms.ac.rs

NENAD VUŠOVIĆ
University of Belgrade, Technical Faculty in Bor, Bor, Serbia; nvusovic@tfbor.bg.ac.rs

MLADEN PEJIĆ
Independent researcher; mladenpejic@outlook.com

SLAVICA MIHAJLOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia; s.mihajlovic@itnms.ac.rs

MILICA VLAHOVIĆ
University of Belgrade, Institute of Chemistry, Technology and Metallurgy-National Institute of the Republic of Serbia, Belgrade, Serbia; m.vlahovic@ihtm.bg.ac.rs; mvlahovic@tmf.bg.ac.rs

Abstract: *Mechanical activation is widely used to induce solid-state reactions; however, distinguishing between structural disordering and true chemical transformation remains a key challenge. This study proposes a combined methodological approach based on X-ray diffraction (XRD) and atomic absorption spectroscopy (AAS) for evaluating the progress of mechanochemical processes in the MgO-TiO_2 system. Mechanical activation was performed in a high-energy vibration mill, and the evolution of the system was monitored at different time intervals. XRD analysis revealed a gradual decrease in the intensity and broadening of diffraction peaks, indicating significant structural disordering and near-complete amorphization of the starting components. However, no new crystalline phases corresponding to magnesium titanate were detected. Complementary AAS results confirmed that the chemical composition of the system remained unchanged throughout the activation process, indicating the absence of a measurable reaction extent. The combined interpretation of structural and chemical data demonstrates that mechanical activation led predominantly to amorphization rather than to synthesis. The results highlight the importance of integrating multiple analytical techniques to reliably assess mechanochemical processes and avoid misinterpretation based on a single method. The proposed approach provides a practical framework for distinguishing between activation-induced structural changes and actual phase formation in solid-state systems.*

Keywords: *mechanical activation; XRD; AAS; amorphization; MgO-TiO_2 system.*

1. INTRODUCTION

Mechanochemical methods have attracted considerable attention as an efficient route for inducing solid-state reactions, particularly in the synthesis of advanced functional materials. Mechanical activation enables the transfer of energy into solid systems through repeated deformation, fracture, and cold-welding processes, leading to the accumulation of structural defects, reduction of crystallite size, and increased reactivity of the treated materials [1]. In recent years, mechanochemistry has emerged as a versatile and sustainable approach for material synthesis, offering reduced solvent usage and improved energy efficiency, and is increasingly recognized as a new frontier in chemical processing [2-4].

Among various solid-state systems, mixtures of alkaline earth metal oxides and titanium dioxide have been extensively studied due to their potential to form technologically important titanates with dielectric, catalytic, and electronic applications [5-9]. In such systems, mechanochemical synthesis is often considered a promising alternative to conventional high-temperature processing routes. Recent advances further demonstrate the applicability of mechanochemical methods in the synthesis of advanced materials, including catalysts and energy-related systems [10,11].

However, despite the theoretical feasibility of these reactions based on thermodynamic considerations, the actual occurrence of mechanochemical synthesis strongly depends on kinetic factors, including the amount of energy introduced into the system, milling conditions, and intrinsic material properties [12-15]. In particular, the

Corresponding author: M. Vlahović;
E-mail: m.vlahovic@ihtm.bg.ac.rs;
DOI: 10.5937/OTEH26121D

MgO-TiO_2 system represents a challenging case, as the formation of $MgTiO_3$ is associated with relatively low driving force compared to other alkaline earth titanates, making the reaction difficult to achieve under purely mechanochemical conditions [16].

A key challenge in the study of mechanochemical processes is the reliable evaluation of reaction progress. Structural characterization techniques, such as X-ray diffraction (XRD), are widely used to monitor phase evolution and crystallographic changes during mechanical activation. However, peak broadening and intensity reduction, commonly observed during high-energy milling, may arise from amorphization and defect formation rather than from the formation of new phases, which can lead to ambiguous interpretation of results.

Complementary chemical analysis methods, such as atomic absorption spectroscopy (AAS), provide quantitative information on the composition of the system and enable estimation of reaction progress. Nevertheless, these techniques do not provide direct insight into structural transformations at the crystallographic level.

Recent studies emphasize the importance of combining multiple analytical approaches in order to obtain a comprehensive understanding of mechanochemical processes and to avoid misinterpretation based on a single technique [17-19].

The aim of this study was to assess the applicability of a combined XRD-AAS approach for evaluating mechanochemical reaction progress and distinguishing activation-induced structural disordering from true mechanochemical synthesis, using the MgO-TiO_2 system as a model case.

2. EXPERIMENTAL

2.1. Materials

Magnesium oxide (MgO) and titanium dioxide (TiO_2) of analytical grade purity (p.a., Merck, Germany) were used as starting materials. The powders were mixed in a stoichiometric ratio corresponding to the formation of $MgTiO_3$.

The initial mixture was prepared using 0.5 mol of each component, resulting in a total mass of approximately 60 g. Prior to mechanical treatment, the powders were homogenized to ensure uniform distribution of components within the reaction system.

2.2 Mechanical activation

Mechanical activation of the MgO-TiO_2 system was performed in a high-energy vibration mill (type MH954/3, KHD Humboldt Wedag, Germany). The milling process was carried out under ambient conditions.

Samples were collected at selected time intervals (60, 180, 330, and 1000 min) in order to monitor the evolution of the system during mechanical activation. The mass of the processed material and milling conditions were adjusted according to the operational characteristics of the mill to ensure efficient energy transfer.

2.3. Characterization methods

X-ray diffraction (XRD)

Structural characterization of the starting materials and mechanically activated samples was performed using X-ray diffraction (XRD) analysis (Philips PW-1700 diffractometer, Netherlands) with Cu Kα radiation. The measurements were carried out over an appropriate 2θ range to identify phase composition and monitor structural changes during activation.

Atomic absorption spectroscopy (AAS)

Chemical analysis of the system was conducted using atomic absorption spectroscopy (AAS) (Analyst 300, PerkinElmer, USA). Prior to analysis, the samples were dissolved in a 10% acetic acid solution.

The obtained concentrations of magnesium and titanium were used to evaluate the progress of the mechanochemical reaction.

Reaction progress evaluation

The mechanochemical reaction progress was estimated based on the results of chemical analysis using a relation derived from the stoichiometry of the MgO-TiO_2 system:

$$S = \left(1 - \frac{0.25\,(c - bc_{Ti})}{c_0 w}\right)100 \qquad (1)$$

where c represents the concentration of magnesium in the solution, c_0 is the initial sample mass, c_{Ti} is the concentration of titanium, w is the mass fraction of magnesium in the initial mixture, and b is the stoichiometric coefficient related to the formation of $MgTiO_3$.

3. RESULTS AND DISCUSSION

3.2 Structural evolution during mechanical activation (XRD analysis)

The structural evolution of the MgO-TiO_2 system during mechanical activation was monitored by X-ray diffraction (XRD) analysis at different milling times. The corresponding diffraction patterns are presented in Figure 1.

The diffraction patterns of the mixture after 60 min of activation show well-defined peaks corresponding to crystalline MgO and TiO_2. With increasing activation time, a gradual decrease in peak intensity and simultaneous peak broadening are observed. These changes indicate progressive reduction in crystallite size and accumulation of lattice defects, which are typical consequences of high-energy mechanical treatment. Such structural evolution is commonly observed during high-energy mechanical activation and is generally attributed to defect accumulation and progressive amorphization of the crystalline phases [17–19].

After prolonged activation (1000 min), the diffraction peaks become significantly broadened and reduced in

intensity, suggesting near-complete amorphization of the system. Importantly, no new diffraction maxima corresponding to magnesium titanate ($MgTiO_3$) or any other reaction products are detected throughout the entire activation period.

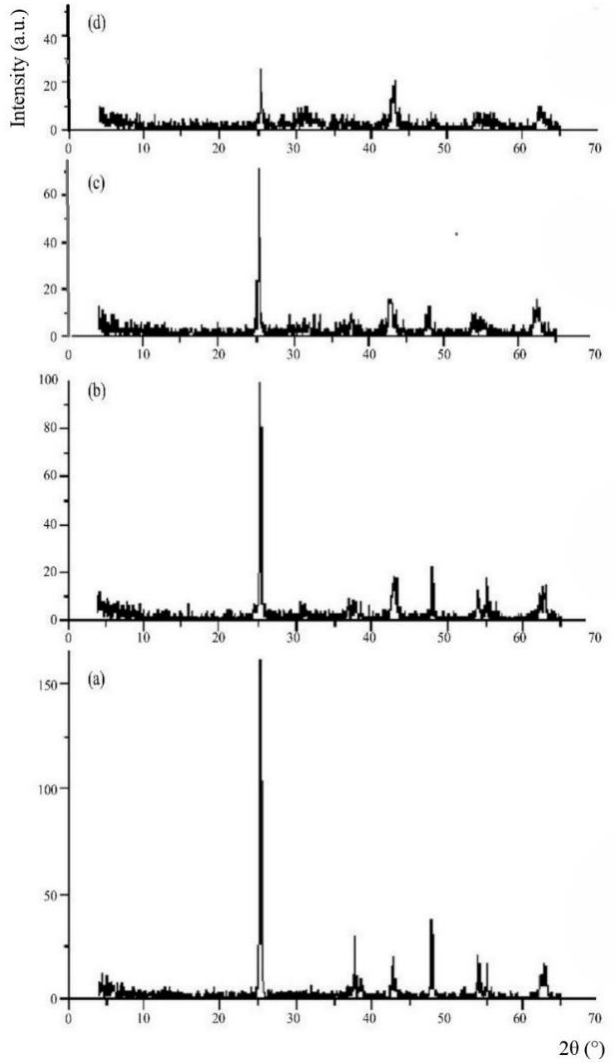


Figure 1. XRD patterns of the MgO-TiO_2 system during mechanical activation:a) 60 min, b) 180 min, c) 330 min, and d) 1000 min

The absence of new crystalline phases, despite intensive structural disordering, indicates that mechanical activation under the applied conditions leads predominantly to structural transformation rather than to the formation of a new compound. Therefore, XRD results clearly demonstrate that the process is governed by amorphization and defect generation, without evidence of mechanochemical synthesis. Similar observations have been reported for oxide systems in which mechanical activation promotes structural disordering without providing sufficient energy for complete solid-state reaction [12,14].

3.2. Chemical analysis and evaluation of reaction progress (AAS)

To complement the structural analysis, the chemical composition of the system during mechanical activation was evaluated using atomic absorption spectroscopy (AAS). The calculated reaction progress as a function of activation time is presented in Figure 2.

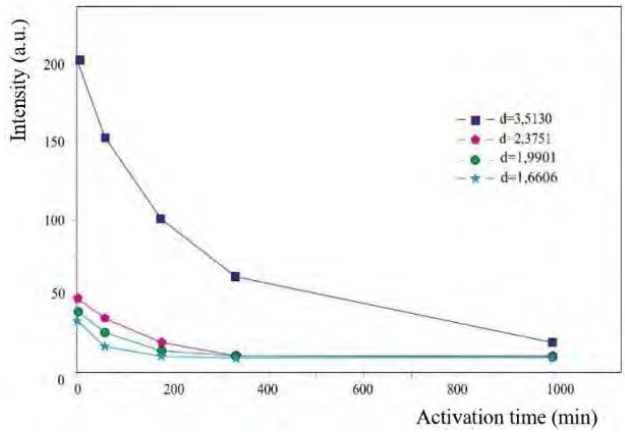


Figure 2. Mechanochemical reaction progress in the MgO-TiO_2 system as a function of activation time, calculated from AAS data

The results show that the concentrations of magnesium and titanium in the analyzed samples remain essentially unchanged throughout the entire duration of mechanical activation. Accordingly, the calculated reaction progress remains negligible, indicating that no measurable chemical transformation occurs within the system. The absence of measurable compositional changes confirms that structural disordering alone should not be interpreted as evidence of mechanochemical reaction, highlighting the importance of complementary chemical analysis.

These findings are consistent with the XRD results, confirming that the observed structural changes are not accompanied by chemical reaction leading to the formation of magnesium titanate. The absence of compositional variation demonstrates that the system remains in its initial chemical state, despite undergoing significant structural disordering.

Thus, AAS analysis provides quantitative confirmation that the mechanochemical reaction progress is effectively zero under the applied experimental conditions.

3.3 Methodological validation of the combined XRD-AAS approach

The obtained results highlight the importance of employing a combined analytical approach for reliable evaluation of mechanochemical processes. When used independently, each method provides only partial insight into the system evolution. This observation is consistent with previous studies emphasizing that reliable interpretation of mechanochemical processes requires complementary analytical techniques [18,19].

XRD analysis alone reveals significant structural changes, such as peak broadening and intensity reduction, which may be misinterpreted as indicators of mechanochemical reaction progress or phase transformation. However, without complementary chemical information, it is not possible to distinguish between amorphization and the actual formation of a new phase.

On the other hand, AAS analysis provides accurate information on the chemical composition and enables quantification of reaction progress, but does not offer insight into structural transformations occurring at the crystallographic level.

The combined application of XRD and AAS overcomes these limitations by integrating structural and chemical perspectives. The combined interpretation of structural and chemical data is schematically presented in Figure 3, illustrating the methodology for distinguishing amorphization from mechanochemical reaction progress.

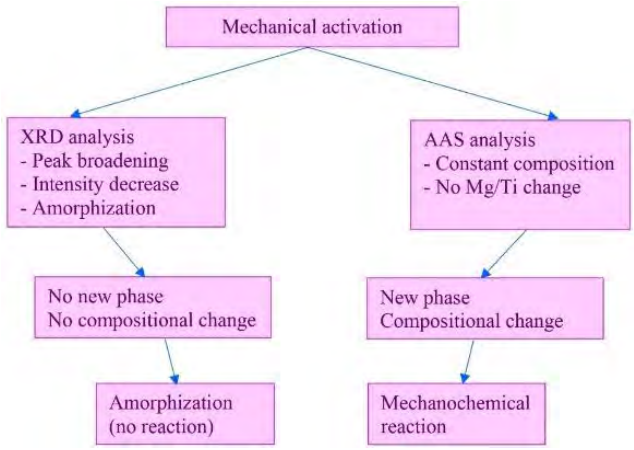


Figure 3. Schematic representation of the combined XRD-AAS methodology for evaluating mechanochemical reaction progress and distinguishing amorphization from mechanochemical synthesis

In addition to the experimentally observed case, the schematic also includes a general scenario in which the combined XRD-AAS approach confirms the occurrence of a mechanochemical reaction, thereby demonstrating the broader applicability of the proposed methodology.

In the present study, this approach clearly demonstrates that intensive mechanical activation leads to pronounced amorphization without any detectable mechanochemical reaction.

The proposed methodology enables reliable discrimination between structural disordering and true mechanochemical reaction progress, thereby reducing the risk of misinterpretation of experimental results. This is particularly important in systems where reaction kinetics are limited or where high-energy input predominantly induces structural defects rather than chemical transformation.

4. CONCLUSION

In this study, the progress of mechanochemical processes in the MgO-TiO_2 system was evaluated using a combined XRD-AAS approach. The results show that intensive mechanical activation leads to significant structural disordering and near-complete amorphization of the starting components, as evidenced by XRD analysis. However, no formation of new crystalline phases was detected.

Complementary AAS analysis confirmed that the chemical composition of the system remained unchanged throughout the activation process, indicating that no measurable mechanochemical reaction occurred.

The combined interpretation of structural and chemical data demonstrates that mechanical activation, under the applied conditions, induces predominantly amorphization rather than mechanochemical synthesis.

The proposed XRD-AAS methodology enables reliable evaluation of mechanochemical reaction progress and provides a practical framework for distinguishing structural disordering from true chemical transformation. This approach reduces the risk of misinterpretation when relying on a single analytical technique and is applicable to a wide range of solid-state systems. The presented approach is particularly valuable in systems with limited reaction kinetics, where distinguishing between activation-induced structural changes and true reaction progress is essential.

Acknowledgement

The authors would like to thank the Ministry of Science, Technological Development and Innovations of the Republic of Serbia for support (Grant Nos: 451-03-33/2026-03/200051, 451-03-33/2026-03/200023, 451-03-34/2026-03/200131, and 451-03-33/2026-03/200026).

References

[1] BOLDYREV,V.,V.: *Mechanochemistry and mechanical activation of solids*, Solid State Ionics, 63-65 (1993) 537-543.

[2] REYNES,J.,F., LEON,F., GARCÍA,F.: *Mechanochemistry for Organic and Inorganic Synthesis*, ACS Organic & Inorganic Au, 4(5) (2024) 432-470.

[3] LI,S., LIU,J., WANG,Z.,L., WEI,D.: *Mechano-driven chemical reactions*, Green Energy & Environment, 10(5) (2025) 937-966. https://doi.org/10.1016/j.gee.2024.08.001

[4] CHETRY,A.,B.: *Mechanochemistry: A new frontier in chemical synthesis*, Journal of Chemical Research, 49(3) (2025) 1-16.

[5] TYLISZCZAK,B., GACA,K., SOBCZAK-KUPIEC,A., DULIAN,P.: *Mechanochemical synthesis and investigations of calcium titanate powders and their acrylic dispersions*, Journal of the European Ceramic Society, 34(10) (2014) 2259-2264.

[6] CHERDCHOM,S., RATTANAPHAN,T., CHANADEE,T.: *Calcium Titanate from food waste: Combustion Synthesis, Sintering, Characterization, and Properties*, Advances in Material Science and Engineering, 2019 (2019) 9639016.

[7] MANAFI,S., JAFARIAN,M.: *Synthesis of perovskite CaTiO3 nanopowders with different morphologies*

by mechanical alloying without heat treatment, International Journal of Physical Science, 8(23) (2023) 1277-1283.

[8] RALPHS,K., HARDACREAND,C., JAMES,S.: *Application of heterogeneous catalysts prepared by mechanochemical synthesis*, Chemical Society Reviews, 42 (2013) 7701-7718.

[9] ĐORĐEVIĆ,N.,G., MATIJAŠEVIĆ,S.,D., VUŠOVIĆ,N.,M., MIHAJLOVIĆ,S.,R., VLAHOVIĆ,M.,M.: *Phase transformations kinetics in barium titanate synthesis by mechanochemical processing*, Hemojska industrija, 79(4) (2025) 219-231.

[10] HU,Y., LI,B., YU,C., FANG,H., LI,Z.: *Mechanochemical preparation of single atom catalysts for versatile catalytic applications: A perspective review*, Materials Today, 63 (2023) 288-312.

[11] ARFELIS,S., MARTÍN-PERALES,A.I., NGUYEN,R., PÉREZ,A., CHERUBIN,I., LEN,C., MALPARTIDA,I., BALA,A., FULLANA-I-PALMER,P.: *Linking mechanochemistry with the green chemistry principles: Review article*, Heliyon, 10(4) (2024) e34655.

[12] NIKZAD,L., GHOFRANI,S., MAJIDIAN,H., EBADZADEH,T.: *Effect of Ball Milling on Reactive Microwave Sintering of MgO-TiO2 System*, ACERP, 2(3) (2016) 25-28.

[13] KHALAJABADI,S., RAFIQ,M., KADIR,A., IZMAN,S., BAKHSHESHI-RAD,H., FARAHANY,S.: *Effect of mechanical alloying on the phase evolution, microstructure and bio-corrosion properties of a Mg/HA/TiO2/MgO nanocomposite*, Ceramics International, 40(10) (2014) 16743-16759.

[14] BHUYAN,R., SAHOO,P., BASANTA,K., SARANGI,A.: *Structural and Thermal Study of Mg_2TiO_4 Nanoparticles Synthesized by Mechanical Alloying Method*, Micro and Nanosystems, 12(2) (2020) 87-91.

[15] YANG,H., ZHIHONG,L., YUMEI,Z.: *Effect of $MgO–TiO_2–SiO_2$ additions on in-situ anisotropic grains growth and mechanical properties of corundum abrasive using pseudo-boehmite as raw material*, Ceramics International, 46 (2) (2020) 1934-1939.

[16] VIDOJKOVIĆ,V. Proučavanje mehanizma i kinetike mehanohemijske sinteze neorganskih soli kod reakcija neutralizacije. Doktorska disertacija, Univerzitet u Beogradu, 2001.

[17] BALÁŽ,P.: *Mechanochemistry in nanoscience and minerals engineering,* 1st edition, Springer Berlin, Heidelberg, 2008.

[18] JAMES,S.,L., ADAMS,C.J., BOLM,C., BRAGA,D., COLLIER,P., FRIŠČIĆ,T., GREPIONI,F., KENNET H.D.M., HYETT,G., JONES,W., KREBS,A., JAMES,M., MAINI,L., ORPEN,A.,G., PARKIN,I.,P., SHEAROUSE,W.,C., STEED,J.,W., WADDELL,D.D.: *Mechanochemistry: opportunities for new and cleaner synthesis*, Chemical Society Reviews, 41(1) (2012) 413-447.

[19] DO,J-L., FRIŠČIĆ,T.: *Mechanochemistry: a force of synthesis*, ACS Central Science, 3(1) (2017) 13-19.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE ANALYSIS OF DIFFERENT 3D PRINTED IGNITER CHARGES ON INTERIOR BALLISTIC PROPERTIES

VESELIN ŽIVANOVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, veselin.zivanovic.best@gmail.com

JELENA MOJSILOVIĆ
University of Defence, Veljka Lukica Kurjaka, 11042, Belgrade, Serbia, jelena.mojsilovic@mod.gov.rs

MARKO RADISAVLJEVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, markoradisavljević.94@gmail.com

VERA JOKSIMOVIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, joksimovic.vera.018@gmail.com

JOVANA MANDIĆ
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, jovana.mandic@mod.gov.rs

IRINA ULANOVA
Military Technical Institute, Ratka Resanovica 1, 11030, Belgrade, Serbia, srb96@mail.ru

Abstract: *This paper aims to show how different designs of igniter charges affect interior ballistic properties. Different geometries were achieved with 3D printed technology. All igniter charges were prepared with an identical number and amount of pyrotechnic compositions to ensure consistency in released chemical energy. The experiments were conducted in experimental rocket motor with fixed and identical free volume as a flight rocket motor. To accurately monitor combustion behavior, operating pressures were measured at two points with the test setup using calibrated pressures transducers. The collected data were analyzed to assess the influence of geometric variations on ignition efficiency, pressure increase time and maximum pressure values.*

Keywords: *igniters, 3D printing technology, rocket motor*

1. INTRODUCTION

With the development of new additive manufacturing technologies, particularly 3D printing [1-2], it has become possible to create components with complex geometries that were previously economically viable only in mass production. Traditional manufacturing methods often required expensive tooling for each geometry, making small-batch or prototype production inefficient. In contrast, 3D printing allows rapid prototyping and cost-effective fabrication of custom parts without the need for molds or tooling, making it especially valuable in research and development environments. Ideas can be quickly transformed into physical components, tested, and iterated with minimal delay.

Additive manufacturing is increasingly utilized in the defense and aerospace industries across the globe [3-4]. In modern warfare conditions, 3D printing enables field-level production of urgently needed components, directly supporting operational readiness and logistical flexibility. Rapid on-site fabrication reduces downtime and increases system reliability in mission-critical applications.

One of the most critical components in a solid rocket propulsion system [5] is the igniter housing. The igniter must ensure precise and consistent initiation of combustion without damaging the propellant grain or nozzle throat [6]. Therefore, the housing must be mechanically robust, thermally stable, and capable of withstanding a wide range of environmental stressors—such as impact, vibration, and thermal cycling—that occur throughout the rocket's operational life cycle [7].

Another essential requirement for igniter housings is hermetic sealing and safe handling. In large-scale production scenarios, igniters are assembled and installed by trained personnel under standardized procedures. Therefore, the housing must not only guarantee performance but also meet stringent safety standards during storage, transport, and integration.

Before finalizing a design for mass production, several iterative prototypes must be tested under identical operating conditions to those of the final application. For this purpose, all igniters in this study were tested in experimental rocket motor with a fixed and controlled internal free volume.

Additive manufacturing enables the use of a wide range of materials with varying densities, thermal tolerances, and mechanical strengths [8-9].

Corresponding author: V.Živanović;
E-mail: veselin.zivanovic.best@gmail.com
DOI: 10.5937/OTEH26122Z

In this work, igniter housings were produced using a Creality K1C 3D printer (Figure 1), capable of printing high-performance polymers reinforced with carbon fiber, and supporting extrusion temperatures up to 300°C. The printer's build volume is 220 × 220 × 250 mm, with a maximum print speed of 600 mm/s, allowing for rapid and scalable prototyping.

Figure 1. Creality K1C 3D printer

2. DESIGN

Two different igniter housing models were designed using SolidWorks. Each housing consisted of three separate parts to allow for easy loading of pyrotechnic compositions and visual inspection before final sealing. The STL models were processed in the Creality slicing software (Figure 2), where part orientation, material selection, layer thickness, and print resolution were configured. Based on the slicing parameters, the estimated print time for a complete igniter housing was approximately 45 minutes. The selected material was 1.75 mm PLA filament, with a temperature resistance of up to 260°C.

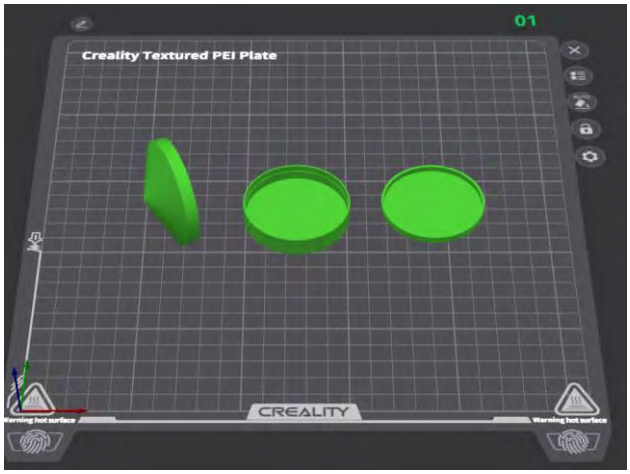


Figure 2. Creality slicing software – geometry 1

Both housing designs featured identical internal free volume to allow consistent loading of pyrotechnic compositions [10]. A total of six igniters were fabricated, with three for each housing geometry. All igniters were hermetically sealed and subjected to dimensional and mass control verification.

The internal architecture of each housing consisted of a primary cylindrical chamber (60 mm diameter) containing the initial ignition mixture—black powder with a pyrotechnic initiator cap and Composition No. 1. To this base structure, a second modular chamber of varying geometry was added. In Geometry 1 (Figure 3), this second chamber took the form of a regular conical shape containing Composition No. 2.

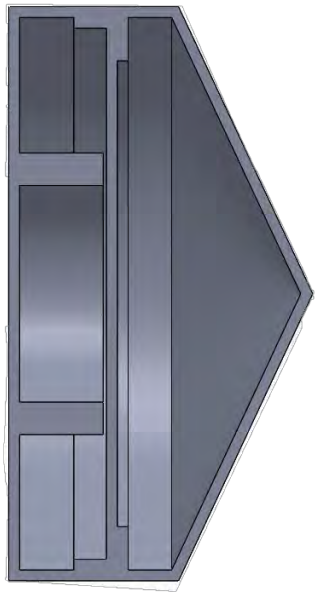


Figure 3. Geometry 1

In Geometry 2 (Figure 4), the second chamber featured an inverted conical cavity with a 60-degree angle, also housing Composition No. 2.

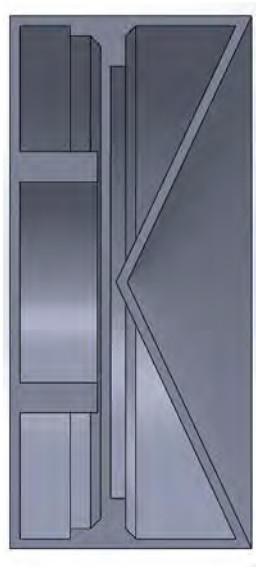


Figure 4. Geometry 2

The wall thickness of all printed housings was 1 mm. The free volume of Chamber 1 was 0.026 dm^3, while Chamber 2 had a volume of 0.042 dm^3.

For the third test configuration, the same pyrotechnic composition was loaded into an ethyl cellulose pouch, without a rigid printed housing structure. This configuration served as a control to observe the ignition behavior of the composition in a less confined environment. The results were not presented in this paper.

The used pressure transducers are designed for measuring static and dynamic liquid and gas pressures. They are available in various measuring ranges, from 1 to 3000 bar. Transducers with the 0 to 3000 bar measuring ranges work according to the strain gage principle. The strain gage application is located on the side of the tube away from the measurement medium in a hermetically sealed reference chamber. This protects it from environmental influences.

3. EXPERIMENT

All igniter charges were tested in experimental rocket motor (Figure 5) equipped with a calibrated pressure transducers P3IC up to 200 bar.

The transducers housings are made from stainless steel and hermetically seal the measurement system to protect it against any harmful external influences, so that even if the operating conditions continue to be rough, the reliability and precision of the transducer are not impaired.

The resulting data were acquired and visualized as pressure-time (P-t) diagrams. Accurate pressure measurement is important in interior ballistic testing, and uncertainty was minimized through calibration, thermal compensation, and signal filtering. Remaining uncertainties were evaluated using standard measurement uncertainty analysis.

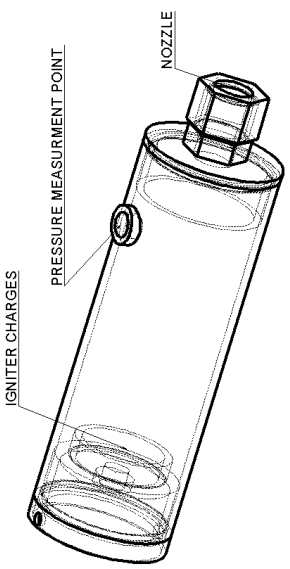


Figure 5. Experimental rocket motor

The pressure-time diagram for Geometry 1 (Figure 6) shows a maximum pressure of approximately 31 bar (y-axis), with moderate variation between trials.

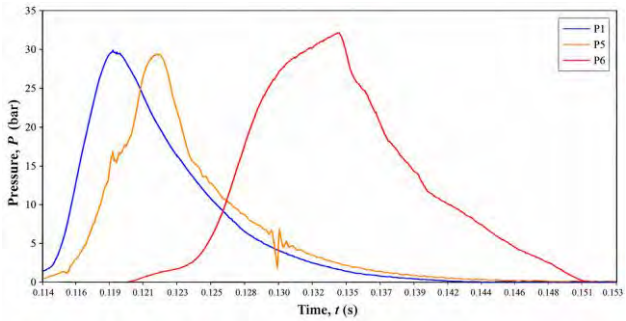


Figure 6. The pressure-time diagram for Geometry 1

The results presented in Table 1 include the maximum pressure and effective operating time.

Table 1. Geometry 1 test results

Igniter #	P	P (bar)	t (s)
#1	P1	30	0.014
#2	P5	29.8	0.013
#3	P6	31.6	0.016

The corresponding diagram for Geometry 2 (Figure 7) reveals a slightly higher maximum pressure—around 10% higher than Geometry 1—under otherwise identical conditions.

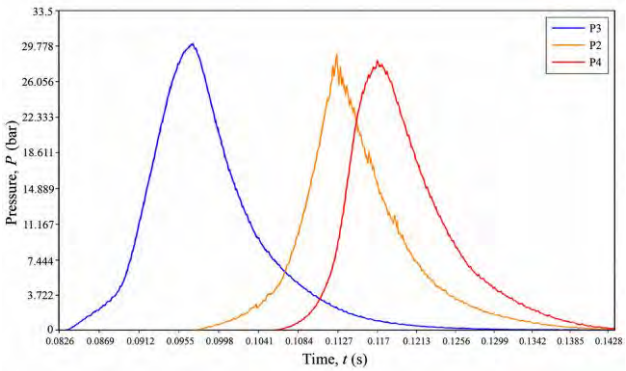


Figure 7. The pressure-time diagram for Geometry 2

The results presented in Table 2 include the maximum pressure and effective operating time.

Table 2. Geometry 1 test results

Igniter #	P	P (bar)	t (s)
#4	P2	27.5	0.014
#5	P3	29.8	0.015
#6	P4	27.3	0.014

These results show that internal geometry, even with identical pyrotechnic mass and free volume, can significantly affect pressure generation and combustion dynamics. The effect is likely related to localized flame propagation, gas flow paths, and the surface-area-to-volume ratio, all of which influence the rate of pressure increase.

Furthermore, the study demonstrates the efficiency and repeatability of 3D printing in the design, fabrication, and testing of igniter housings. Additive manufacturing offers a rapid and cost-effective method for prototyping and even low-rate production of components with complex internal geometries. It enables functional validation before mass production, reduces lead times, and allows for immediate modifications.

3. CONCLUSION

The results of this study confirm that internal geometry of igniter housing has a significant role in effect on interior ballistic behavior, even when the pyrotechnic composition and free volume remain constant. Among the tested configurations, Geometry 2—featuring a recessed conical cavity—demonstrated a ~10% increase in maximum pressure compared to Geometry 1. While the overall impulse remained comparable across designs, the rate of pressure increases and flame propagation dynamics were noticeably affected by structural differences.

These findings validate the hypothesis that optimized igniter geometry can enhance ignition efficiency and combustion control in solid rocket motors. Moreover, the successful application of 3D printing technology for rapid prototyping of complex igniter housings highlights its practical value in both research and low-volume production environments. The use of carbon-fiber-reinforced and high-temperature-capable filaments provides an adequate balance of mechanical strength, thermal stability, and manufacturing speed.

Future work will focus on numerical simulations (finite element methods or CFD analysis) of thermal and gas-dynamic processes within the igniter chamber to correlate experimental findings with predictive models. Additionally, further optimization of geometry, material selection, and surface finish could lead to improved safety, repeatability, and ignition performance under operational conditions.

Acknowledgments

This work was supported by the Ministry of Science, Technological Development and Innovations (Serbia), Contract No. 451-03-34/2026-03/200325 and University of Defense, Military Academy, Proj. No. VA-TT/1/22-24.

References

[1] DUDA T., RAGHAVAN L.V.: 3D Metal Printing Technology, IFAC - PapersOnLine 2016, 49, 103–110. [*https://doi.org/10.1016/j.ifacol.2016.11.111]*

[2] CHEN Y.P., YANG M.D.: Micro-Scale Manufacture of 3D Printing, Appl. Mech. Mater, 2014, 670, 936–941.[*https://doi.org/10.4028/www.scientific.net/AMM.670-671.936]*

[3] R. Marchan, A. Oleshchenko, S. Vekilov, M. Arsenuk, and O. Bobrov, "3D printed acoustic igniter of oxygen-kerosene mixtures for aerospace applications".

[4] "NASA rocket engine igniter made of Inconel and copper by hybrid 3D laser printing," Shape Memory and Superelastic Technologies. Accessed: Jul. 11, 2026.[Online].Available:https://www.asminternational.org/smst/results/-/journal_content/56/10192/27623036/NEWS/

[5] GLIGORIJEVIĆ N., DIVAC S., LEMIĆ D., PETRIĆ A., SAVKOVIĆ M., ČOLAKOVIĆ M., VUČUROVIĆ O., JOVANOVIĆ V.: Rocket Motors with Solid Propellant, Military Technical Institute Belgrade, Serbia, 2013, ISBN 978-86-81123- 63-8.

[6] GLIGORIJEVIĆ N., NEŠIĆ J., BOGDANOV J., PAVKOVIĆ B., ŽIVKOVIĆ S., TANASKOVSKI B.: *Influence of the Ignition Mixture Particle Size on Rocket Motor Igniter Pressure Gradient,* Scientific Technical Review, 2018, Vol. 68, No. 1, pp. 33-39.

[7] G. B. Pulpea, A. Voicu, L. Matache, L. Haller, A. D. Mandache-Dodoiu, and R. Ginghină, "Numerical Simulation of Interior Ballistics for Pyrotechnics Systems," no. 1, 2017.

[8] M. F. Gomes, "Internal Ballistics Simulation of a Solid Propellant Rocket Motor".

[9] S. F. Herano, "A thesis submitted In Partial Fulfillment of the Requirements for the," vol. 9, no. 8, 2021.

[10] KRSTOVIĆ M., STANČIĆ T., NESTOROVIĆ K., BAJIĆ D., BOGDANOV J., BAJIĆ Z.: *Vacuum Stability Evaluation of Pyrotechnic Mixture – Double Base Propellant Compatibility,* 28th Consecutive Seminar on New Trends in Research of Energetic Materials NTREM, Chezch Republic, Pardubice, 22-24 April 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

THE EFFECT OF PALLADIUM LOADING ON PD-MODIFIED SB CATALYSTS IN THE ELECTROOXIDATION OF ETHANOL

MILICA G. KOŠEVIĆ*
Department of Electrochemistry, Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, Belgrade, Serbia, milica.kosevic@ihtm.bg.ac.rs

NEBOJŠA D. NIKOLIĆ
Department of Electrochemistry, Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, Belgrade, Serbia, nnikolic@ihtm.bg.ac.rs

JELENA D. LOVIĆ
Department of Electrochemistry, Institute of Chemistry, Technology and Metallurgy, University of Belgrade, Njegoševa 12, Belgrade, Serbia, jelena.lovic@ihtm.bg.ac.rs

Abstract: *The ethanol electrooxidation reaction is a key process in direct ethanol fuel cells. However, its practical application is hindered by catalyst poisoning and limited stability. In this work, Pd–Sb electrocatalysts (Pd@Sb/GC) were synthesized via a two-step electrodeposition route. The Pd/Sb ratio was fixed at 80:20, while the total metal loading was systematically varied. The catalytic activity toward ethanol oxidation in alkaline media was evaluated using cyclic voltammetry and chronoamperometry. All samples exhibited the characteristic dual-peak behavior associated with ethanol oxidation and the removal of adsorbed intermediates. The Pd@Sb/GC catalyst with a metal loading of 4.41 µg Pd demonstrated the highest activity, achieving a peak current density of 600 mA mg^{-1}. In addition to the highest current density, it showed a more negative onset potential, indicating enhanced catalytic performance. This improvement is attributed to the formation of well-dispersed Pd islands. At higher Pd loadings, increased agglomeration led to a decrease in mass activity. The Pd@Sb/GC (4.41 µg Pd) also exhibited the highest jf/jb ratio, close to 2, along with a smaller peak separation, indicating improved resistance to poisoning and enhanced reaction reversibility.*

Keywords: *Palladium-based catalyst, catalyst poisoning tolerance, surface morphology effect, electrocatalytic stability*

1. INTRODUCTION

Ethanol is primarily utilized as a fuel in direct ethanol fuel cells (DEFCs) [1]. This growing interest stems from the urgent global demand to transition away from fossil-fuel-dependent energy systems. The high theoretical energy density of ethanol (8.0 kWh kg−1), combined with its low toxicity and ease of storage as a liquid, makes it an ideal candidate for sustainable power generation. These specific advantages give it a distinct edge over alternative fuels like methanol or compressed hydrogen. Besides that ethanol is applied in proton exchange membrane (PEM) electrolyzers to produce hydrogen at lower cell voltages (<1 V) compared to water electrolysis (> 1.8 V), often yielding high-value chemical byproducts like acetic acid [2]. This approach not only reduces the energy footprint of green hydrogen production but also integrates chemical synthesis into the electrochemical cycle. Consequently, a highly favorable dual benefit of clean energy production and chemical manufacturing is realized simultaneously. Likewise, well known are ethanol biosensors for amperometric determination of ethanol in alkaline media with the aid of enzymes (e.g., Alcohol Oxidase - AOX) or Ni-based catalysts (e.g., Ni-Schiff base complexes), measuring the generated current to determine concentration in applications like breathalyzers [3]. The development of non-enzymatic sensors, in particular, has gained traction due to their superior thermal stability and longer shelf life compared to biological recognition elements. This makes them ideal for demanding point-of-care testing and long-term industrial monitoring.

Ethanol is a key organic molecule in electrochemistry, and Pd is acknowledged as one of the most effective metal for the ethanol electrooxidation reaction (EOR) [1, 4]. In alkaline environments, Pd exhibits faster kinetics for C–H bond activation and higher CO-tolerance than platinum-based benchmarks. However, several key factors hinder its practical use, primarily related to catalytic deactivation, inefficient reaction pathways, and stability issues. The accumulation of strongly adsorbed carbonaceous intermediates, such as COads and CHx,ads, often blocks the active Pd sites and leads to a rapid decay in current density. As a result, a dramatic loss in overall cell efficiency occurs over extended operating periods. To address these challenges, research is focused on alloying Pd with other metals (e.g., PdAg, PdNi, PdAu, PdSn) to modify the electronic structure, improve C–C bond cleavage, and promote the formation of OH– species for better CO oxidation [1, 4]. These promoters typically

Corresponding author: M. Košević;
E-mail: milica.kosevic@ihtm.bg.ac.rs
DOI: 10.5937/OTEH26123K

operate via the bifunctional mechanism, where the second metal provides oxygen species at lower potentials to facilitate the oxidative removal of poisons from the Pd surface. This cooperative interaction continually regenerates the active metal sites. While transition metals have been extensively studied as promoters, the exploration of main-group metalloids like antimony (Sb) remains limited. This is an oversight, as antimony has a unique potential to favorably tune the electronic states of noble metals while altering surface oxophilicity.

In this work, a bimetallic Pd-on-Sb catalysts were synthesized by electrodeposition of Pd on pre-prepared Sb nanoparticles using chronopotentiometry. This sequential deposition strategy was employed to create a specific surface architecture where Pd "islands" are anchored onto Sb sites. This arrangement maximizes the synergistic interaction between the two metals. This specific architecture ensures that a high number of active boundary zones are exposed. The electrocatalytic behavior of Pd-on-Sb catalysts, with different amount of electrodeposited Pd and Sb and constant Pd/Sb ratio, was systematically investigated toward EOR using cyclic voltammetry and chronoamperometry, allowing for a clear comparison with the performance of pure Pd. Special attention was paid to the effect of total metal loading on the mass activity and the jf/jb ratio, providing insights into how catalyst dispersion influences poisoning resistance. Ultimately, this approach provided new insights into how the spatial arrangement of the Pd-Sb interface influences reaction reversibility.

2. EXPERIMENTAL

Preparation and characterization of the catalysts

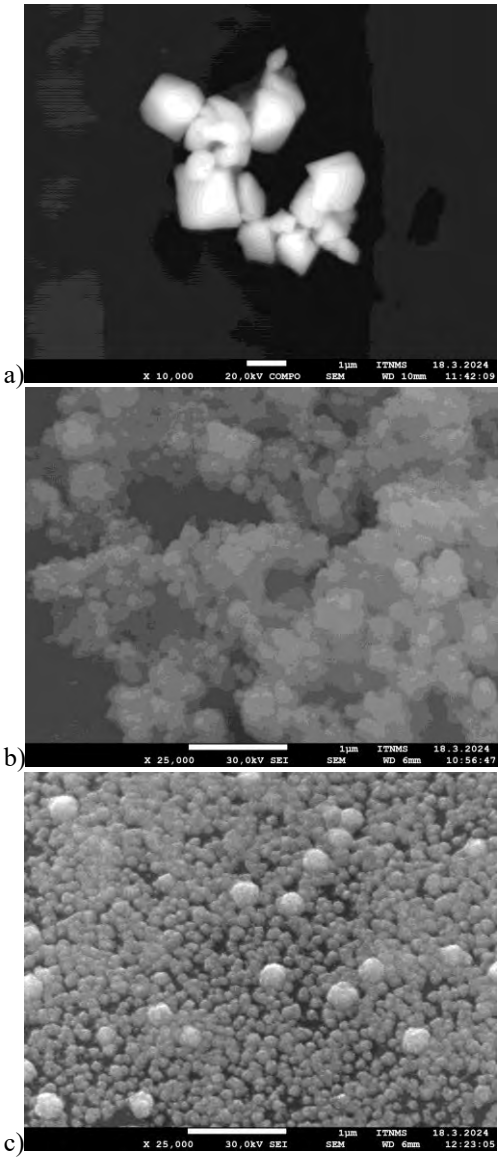
The Pd-on-Sb (designated as Pd@Sb/GC) catalysts were synthesized using a two-step electrodeposition method as was previously described [5]. The Pd@Sb/GC catalysts with a constant atomic ratio of 20 at.% Sb-80 at.% Pd was obtained by controlling the deposited charge of Sb (3, 6, and 12 mC) and Pd (8, 16, and 32 mC). The electrocatalysts are designated as Pd1@Sb/GC, Pd2@Sb/GC and Pd3@Sb/GC, respectively. For accurate comparison, a Pd electrode was constructed utilizing the same deposition parameters corresponding to one for bimetallic electrode. The current densities were normalized by the mass of Pd, and the calculated metal loadings were 4.41 µg, 8.82 µg, and 17.64 µg for Pd1@Sb/GC, Pd2@Sb/GC and Pd3@Sb/GC, respectively.

Morphologies of electrodeposited Sb, Pd and Sb–Pd catalysts were characterized by the technique of scanning electron microscopy (SEM) – model JEOL JSM-7001F.
Electrochemical experiments were in a standard three compartment electrochemical glass cell with Pt wire as the counter electrode and Ag/AgCl/3.5 M KCl as the reference electrode.

3. RESULTS AND DISCUSSION

3.1 Characterization of the Sb, Pd and Sb–Pd catalysts by the SEM technique

Morphology of Sb, Pd and Sb–Pd obtained by galvanostatic regime of the electrodeposition is shown in Fig 1.



Morphology of a) Sb, b) Pd and c) Sb–Pd catalysts

3.2 Ethanol oxidation reaction on the Pd@Sb catalysts

Electrochemical examination of electrodeposited catalysts in EOR was started by CV measurements. Figure 1 shows the voltammetric responses of Pd@Sb/GC catalysts in 1 M NaOH + 1 M C_2H_5OH. A summary of key CV features related to ethanol oxidation revealed two anodic peaks. The forward scan peak is ascribed to the oxidation of chemisorbed ethanol species. Due to the formation of the oxide layer at higher potentials the electrocatalytic activity decrease. The electrode reactivates during the reverse scan after the oxide is removed from its surface, resulting in a second oxidation peak. The backward scan

peak indicates the removal of carbonaceous species that can form on the Pd catalyst surface but remain incompletely oxidized during the forward scan [6].

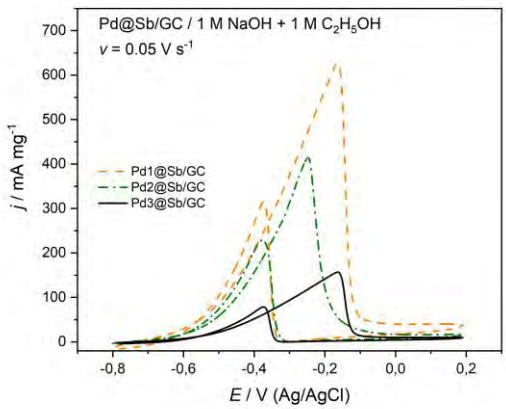


Fig. 2. CVs for the oxidation of 1 M C_2H_5OH recorded on Pd@Sb/GC catalysts in 1 M NaOH. Scan rate: 0.05 V s^{-1}.

According to Fig. 2, the most active Pd@Sb/GC catalyst toward EOR is Pd1@Sb/GC, showing the highest reaction currents in forward and backward direction, along with the shift of the initial potential of the reaction toward more negative values. The observed behavior is explained by the different surface morphology i.e. size effect of Pd islands [7]. An increase in Pd loading resulted in a subsequent decrease in mass activity as the palladium surface changed from isolated Pd nanoparticles to Pd agglomerates. It can be said that the catalytic activity is not directly proportional to surface area, as a higher surface area does not necessarily provide more active sites.

The ratio of the forward anodic peak current density (jf) to the backward anodic peak current density (jb), as well as potential peak separation between the forward and backward scans are crucial indicators of the susceptibility to carbonaceous intermediates that accumulate on the catalyst surface during the EOR. The j_f/j_b ratio of Pd@Sb/GC is measured to be 2, 1.8 and 1.72 for Pd1@Sb/GC, Pd2@Sb/GC, and Pd3@Sb/GC, respectively. The potential peak separation increase in order Pd2@Sb/GC<Pd1@Sb/GC< Pd3@Sb/GC. Reduced peak separation and higher j_f/j_b ratio signifies that the oxidation of intermediate species during the reverse scan occurs closer to the potential of the initial oxidation, indicating better efficiency and less irreversible poisoning. The obtained results are in accordance with chronoamperometric measurements (Fig. 3).

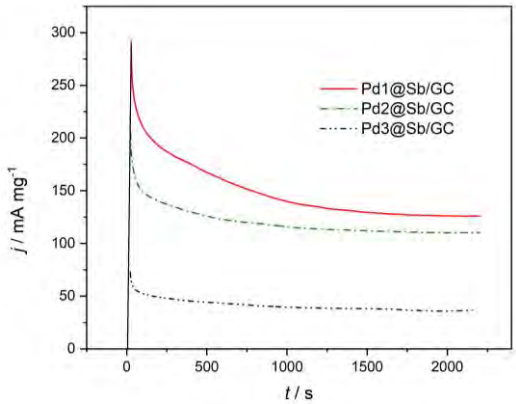


Fig 3. Chronoaperometric curves obtained at E = – 0.35 V for the oxidation of 1 M C_2H_5OH on Pd@Sb/GC.

4. CONCLUSION

The catalytic activity towards the EOR was examined on a catalyst system constructed by the electrochemical deposition of Sb on GC followed by the electrochemical deposition of Pd. According to electrochemical measurements, catalytic performance is heavily dependent on surface morphology rather than total surface area alone. Among the tested samples, the highest performance was delivered by the catalyst with a lower metal loading (4.41 µg Pd, designated as Pd1@Sb/GC). A peak current density of 600 mA mg−1 was achieved by this sample, alongside a significantly more negative onset potential. This enhanced activity is driven by the formation of well-dispersed, isolated Pd islands. Conversely, nanoparticle agglomeration was triggered by higher metal loadings, causing mass activity to be severely restricted.

Beyond high current densities, stable operation was demonstrated by the optimized Pd1@Sb/GC catalyst. The highest jf/jb ratio of approximately 2 and a tightly reduced potential peak separation were yielded. This proves that highly reversible reactions are undergone by the material and a strong resistance to carbonaceous poisoning is possessed. Ultimately, the vital importance of controlling mass loading in bimetallic designs is underscored by these findings, allowing this optimized Pd–Sb configuration to be positioned as a viable anode candidate for next-generation direct ethanol fuel cells.

Acknowledgment

This research has been financially supported by the Ministry of Science, Technological Development and Innovation of Republic of Serbia (Contract No. 451-03-33/2026-03/2000).

References

[1] BISHNOI,P., MISHRA,K., SIWAL,S.S., GUPTA,V.G., THAKUR,V.K.: Direct Ethanol Fuel Cell for Clean Electric Energy: Unravelling the Role of Electrode Materials for a Sustainable Future, Advanced Energy and Sustainability Research, 5(6) (2024) 2300266.

[2] LO VECCHIO,C., MOSCA,E., TROCINO,S., BAGLIO,V.: Investigation of an Ethanol Electroreforming Cell Based on a Pt1Ru1/C Catalyst at the Anode, Catalysts, 14(7) (2024) 415

[3] GORIUSHKINA,T.B., SOLDATKIN,A.P., DZYADEVYCH,S.V.: Application of Amperometric Biosensors for Analysis of Ethanol, Glucose, and Lactate in Wine, Journal of Agricultural and Food Chemistry, 57(15) (2009) 6528–35

[4] ALMEIDA,C.V.S., EGUILUZ,K.I.B., SALAZAR-BANDA,G.R.: Superior ethanol electrooxidation activity of Pd supported on Ni(OH)2/C. The effect of Ni(OH)2 nanosheets content, Journal of Electroanalytical Chemistry, 878 (2020) 114683

[5] KOŠEVIĆ,M.G., NIKOLIĆ,N.D., RAKOČEVIĆ,L.Z., LOVIĆ,J.D.: Ethanol Oxidation Reaction on Electrodeposited Pd and Sb–Pd Catalysts in Alkaline Solution Containing Na or Li Cations, Journal of The Electrochemical Society, 171(9) (2024) 096512

[6] HASAN,M., NEWCOMB,S.B., ROHAN,J.F., RAZEEB,K.M.: Ni nanowire supported 3D flower-like Pd nanostructures as an efficient electrocatalyst for electrooxidation of ethanol in alkaline media, Journal of Power Sources, 218 (2012) 148-156

[7] SANTOVEÑA-URIBE,A., MAYA-CORNEJO,J., ESTEVEZ,M., SANTAMARIA-HOLEK,I.: Thermodynamic Analysis of Size-Dependent Surface Energy in Pd Nanoparticles for Enhanced Alkaline Ethanol Electro-Oxidation, Nanomaterials, 14(23) (2024) 1966

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

BULLETPROOF/BULLET RESISTANT CERAMIC MATERIALS FOR ARMOR APPLICATIONS

TATJANA VOLKOV-HUSOVIĆ
[3]University of Belgrade, Faculty of Technology and Metallurgy, Belgrade, Serbia, tatjana@tmf.bg.ac.rs

ANA ALIL, SANJA MARTINOVIĆ
[2]University of Belgrade, Institute of Chemistry, Technology and Metallurgy- National Institute of the Republic of Serbia, Belgrade, Serbia, ana.alil@ihtm.bg.ac.rs s.martinovic@ihtm.bg.ac.rs

Abstract: *Ceramics are essential for lightweight armor systems, utilizing high hardness and compressive strength to shatter projectiles and absorb impact energy. Key materials include boron carbide (B_4C), silicon carbide (SiC), and alumina (Al_2O_3), commonly used in personal body armor and vehicle shielding to counter high-velocity ballistic threats. Bulletproof materials, more accurately termed bullet-resistant or ballistic materials, are specialized materials designed to stop or significantly decrease the penetration of munitions. They function by absorbing, dispersing, or breaking up the bullet's kinetic energy through layered, high-strength structures like Kevlar, ceramic, steel, or polycarbonate, rather than being completely impervious. These materials work by either deforming the bullet upon impact with a rigid surface (e.g., ceramics, metals) or capturing the projectile through a dense, flexible weave (e.g., aramid fibers) that distributes energy. In this paper, the most commonly used ceramic materials for armor applications will be discussed; these are: alumina (Al_2O_3), silicon carbide (SiC), and boron carbide (B_4C). Armor functions related to ceramic materials, as well as ceramic applications, will be presented, along with performance considerations related to their strengths and weaknesses.*

Keywords: *bulletproof material, ceramic materials, Boron Carbide (B_4C), Silicon Carbide (SiC), Alumina (Al_2O_3)*

1. INTRODUCTION

Materials for military applications are highly specialized materials engineered for extreme conditions and performance, such as durability and lightweight strength, designed to withstand combat, high temperatures, and harsh environments. Key materials include different materials, including metals, ceramics, and composite materials. Some of the most used belongs to the groups of: titanium and aluminum alloys for aerospace, carbon fiber composites, high-performance ceramics for armor, tungsten, and critical rare earth elements for electronics and propulsion [1,2].

One of the interesting applications of materials, or military application are materials with high resistance to impact, "Bullet-Resistant" or "Bulletproof" materials. While commonly called "bulletproof," most materials are only *bullet-resistant*, meaning their effectiveness depends on the bullet type, caliber, velocity, and distance, often categorized by protection levels [3-5].

Compared with traditional steel bulletproof materials, ceramic materials are lighter in weight and have superior protective performance. They have gradually become the first choice material for lightweight protection.

Most used bulletproof ceramic materials are:

- Alumina Ceramic (Al_2O_3)
- Silicon Carbide Ceramics (SiC)
- Boron Carbide Ceramics (B_4C)

In this paper, the focus will be on the main characteristics of the ceramics, their strengths and weaknesses as bullet-resistant materials, and possible applications.

During ballistic impact, extremely high strain rates occur at the point of contact, while strain rates decrease with the distance from the impact site (Figure 1). In these lower-strain regions, quasi-static fracture behavior becomes relevant and can provide insight into failure mechanisms [3-12]. Generally, materials exhibit more brittle behavior at higher strain rates.

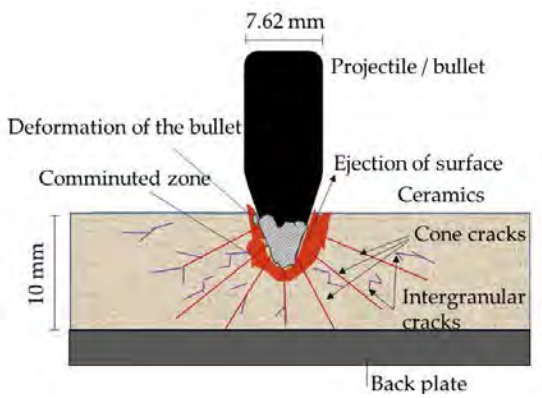


Fig.1. Deformation of the bullet and ceramic tile during ballistic impact. [13]

Relationships between material properties, microstructure, processing conditions, failure mechanisms, and ballistic

Corresponding author: T.Volkov-Husović;
E-mail: tatjana@tmf.bg.ac.rs
DOI: 10.5937/OTEH26124V

performance in order to provide critical discussion of the reasons for different ballistic behavior of selected ceramics [14-19].

- **Material Properties → Microstructure**

Density, hardness, and fracture toughness are direct outcomes of the ceramic's grain size, porosity, and phase composition.
Example: Boron carbide's extreme hardness comes from its crystal structure, but its brittleness is linked to localized amorphization under stress.

- **Microstructure → Processing Conditions**

Hot pressing, sintering temperature, and additives control grain boundaries and porosity.
Alumina processed at lower purity has larger pores, reducing ballistic efficiency compared to high-purity alumina.

- **Processing Conditions → Failure Mechanisms**

Improper sintering can leave residual stresses or weak grain boundaries, leading to crack propagation.
SiC produced via liquid-phase sintering achieves higher flexural strength, delaying catastrophic fracture.

- **Failure Mechanisms → Ballistic Performance**

Ceramics defeat projectiles by shattering or eroding them, but cracks and fragmentation reduce multi-hit capability.
Alumina: fails by brittle fracture, but cost-effective.
SiC: resists armor-piercing rounds but cracks after repeated hits.
B_4C: amorphization reduces effectiveness after localized impact.

2. MATERIALS

2.1. Alumina Ceramic(Al_2O_3)

Alumina provides the best cost-benefit ratio among advanced ceramics, due to the excellent mechanical properties, such as high modulus of elasticity, high refractoriness, high hardness, and relatively lower cost compared to other advanced ceramics. Weaknesses of the alumina-based ceramics are related to the low fracture toughness and low flexural strength. Due to the mechanical properties, the ballistic performance of alumina is lower when compared to SiC and B_4C [8]. The ballistic properties of alumina could be improved by using zirconia or by manufacturing CMCs. This approach could increase fracture toughness and flexural strength by introducing tetragonal zirconia particles or ceramic fibers [9,10].

Some of the relevant mechanical properties of alumina are given in Table 1.

Table 1. Mechanical properties of alumina [7-11]

Properties	Value
Density	3.6-3.95 g/cm³
Flexural strength	200-400 MPa
Hardness	HRA90
Fracture toughness	3.0-4.5 MPa·m$^{1/2}$

Advantages and Properties

Cost-Effectiveness: More economical than alternatives like silicon carbide or boron carbide.
Hardness: High hardness (HRA 90) enables breaking, fracturing, and eroding projectiles.
Durability: Offers excellent chemical stability, corrosion resistance, and heat dissipation, making it suitable for harsh environments.
Performance: High-purity (99%) alumina offers better weight reduction and reliable ballistic results compared to lower purity options.
Key Application Areas include [6,7,11,12]:
Personal Body Armor**:** Used as rigid inserts (plates) in vests for military personnel and police forces to protect against rifle fire.
Vehicle Armor**:** Applied as armored panels in light military vehicles, armored cars, and aircraft, often designed with concave/convex edges to improve joints.
Structural Protection**:** Utilized in security shielding for vital facilities or VIP vehicles.
Composite Structures**:** Frequently bonded with backing materials like fiber-reinforced polymers or UHMWPE (ultra-high molecular weight polyethylene) to increase strength and reduce weight

2.2. Silicon carbide ceramics (SiC)

Silicon carbide ceramics are currently the highest level of bulletproof ceramics. Its special crystal structure forms a diamond-like tetrahedral structure, which gives it extremely high hardness and strong mechanical properties.

Silicon carbide is second only to diamond and boron carbide in hardness and is very light, with a density about 18% lower than that of alumina. Silicon carbide ceramics are mainly used in high-end military equipment, such as special forces bulletproof equipment and key protective parts of armored vehicles.

Table 2. Mechanical properties of silicon carbide ceramics (SiC) [6,13]

Properties	Value
Density	3.10-3.15 g/cm³
Flexural strength	400-730 MPa
Hardness	≥HRA92
Fracture toughness	5.0-5.5 MPa·m$^{1/2}$

Key Performance Benefits

High Hardness: SiC is one of the hardest known materials, second only to diamond and boron carbide. This extreme hardness allows it to erode or "shatter" the hardened steel cores of armor-piercing bullets upon impact.

Lightweight Efficiency: It has a density approximately 18% lower than alumina and 2.5 times less than steel. This reduction in weight is critical for maintaining user mobility without sacrificing protection.

Protection Levels: SiC plates are typically used to achieve NIJ Level III and IV protection standards,

meaning they can stop high-velocity rifle rounds, including .30-06 M2 armor-piercing (AP) threats.

- Thermal & Chemical Stability: The material remains stable at temperatures up to 1600°C/(depending on atmosphere (oxidizing vs. inert), since oxidation can degrade performance) and is highly resistant to corrosion and wear.

The performance of ceramic armor materials depends not only on intrinsic properties but also on processing conditions, microstructure, test configuration, and environment. Similarly, NIJ Levels III and IV describe the ballistic rating of complete armor systems, not the ceramic alone; SiC or B_4C must be integrated with backing materials to achieve these standards. Ballistic efficiency is therefore a system-level outcome, influenced by sintering method, grain size, porosity, and environmental exposure, rather than a single material property.

Ballistic Performance Features:

- **Shock Resistance:** Excellent thermal shock resistance and thermal conductivity prevent premature failure during high-energy impacts.
- **Phase Composition:** Often produced via hot pressing or liquid-phase sintering to optimize microstructure for superior flexural strength.
- **Applications:** Widely utilized in high-end military equipment, such as personal body armor and vehicle armor, owing to its combination of low density and high destructive energy capability.

Critical Limitations

- **Brittleness:** Like all ceramics, SiC is brittle. While it effectively stops initial impacts, it is prone to cracking and may lose effectiveness if subjected to multiple closely spaced hits or "multi-hit" scenarios without specialized backing.
- **Cost:** It is more expensive to produce than alumina due to the complex sintering processes required.
- **Brittle Fracture:** It cannot withstand heavy, sustained gunfire as effectively as some composite or metal-backed systems might, as the entire plate can fail after its structural integrity is compromised.

2.3. Boron Carbide Ceramics (B_4C)

Boron carbide ceramics are known as the “lightest bulletproof ceramics”. Its density is only 2.45-2.52 g/cm³ (about 20% lighter than SiC). $B_{11}C$ has a unique rhombus crystal structure. C-B-C is connected by covalent bonds and has a very stable spatial structure. It is also one of the lightest known superhard materials [14-16].

Boron carbide is mainly used in situations where weight and protection are high, such as air force pilot seat protection, special forces lightweight bulletproof equipment, etc. In addition, it also has good neutron absorption ability and can be used for nuclear protection.

Table 3. Mechanical properties of boron carbide ceramics [11-14]

Properties	Value
Density	2.45-2.52 g/cm³
Flexural strength	200-500 MPa
Hardness	28-35 GPa
Fracture toughness	2.0-4.7 MPa·m$^{1/2}$

Key Advantages in Ballistic Armor

Lowest Density: It is the lightest bulletproof ceramic, with a density of ~2.52 g/cm³. It is roughly 20% lighter than silicon carbide (SiC) and 40% lighter than alumina.

Extreme Hardness: Its high Vickers hardness (~28-35 GPa) allows it to effectively shatter high-velocity armor-piercing (AP) projectiles on impact.

Rifle-Rated Protection: It is a primary component in NIJ Level IV plates, designed to stop .30-06 Springfield M2 armor-piercing rounds.

Weight Sensitivity: Because of its lightness, it is the preferred choice for elite military gear, air force pilot seat protection, and specialized lightweight vests where mobility is critical.

Limitations & Trade-offs

Brittleness & Multi-Hit Performance: It is more brittle than other ceramics. Upon impact, it can suffer localized "amorphization," where its crystal structure breaks down, potentially reducing its effectiveness against subsequent hits in the same area.

Higher Cost: Due to complex manufacturing requirements-needing temperatures above 2000°C for sintering-it is significantly more expensive than alumina or silicon carbide.

Composite Requirement: It is rarely used alone. It is bonded to a ductile backing material (like Kevlar or UHMWPE) to catch the shattered fragments of the ceramic and the bullet.

3. COMPARISON OF THE SELECTED CERAMICS

Some of the relevant properties are listed in Table 4. Representing strengths and weaknesses in comparison of selected ceramics for bulletproof applications.

The ballistic performance of ceramics is not only determined by their intrinsic properties (density, hardness, fracture toughness) but also by their microstructure and processing conditions. For example, high-purity alumina with fine grains exhibits improved hardness and reduced porosity, enhancing its ability to erode projectiles. Silicon carbide, when hot-pressed or liquid-phase sintered, achieves superior flexural strength due to optimized grain boundaries, delaying brittle fracture. Boron carbide’s extreme hardness allows it to shatter armor-piercing rounds, but localized amorphization of its crystal structure under impact reduces multi-hit capability. These failure mechanisms-cone cracking, fragmentation, and amorphization-explain why ceramics must be paired with

backing materials (e.g., UHMWPE, Kevlar) to absorb residual energy and maintain structural integrity. Thus, processing and microstructural control are as critical as the inherent properties in defining the real-world ballistic efficiency of ceramic armor.

Table 4. Comparison of the selected ceramics [11-14, 19-22]

Properties	Value		
	Alumina (Al_2O_3)	Silicon carbide (SiC)	Boron carbide (B_4C)
Density	3.6-3.95 g/cm³	3.10-3.15 g/cm³	2.45-2.52 g/cm³
Flexural strength	200-400 MPa	400-730 MPa	200-500 MPa
Hardness	HRA90	≥HRA92	29-35 GPa
Fracture toughness	3.0-4.5 $MPa \cdot m^{1/2}$	5.0-5.5 $MPa \cdot m^{1/2}$	2.0-4.7 $MPa \cdot m^{1/2}$
Relative cost	medium	low	high

Ceramic armor protects by breaking and eroding projectiles through extreme hardness, while simultaneously dissipating impact energy via fracture and fragmentation. Upon impact, a compressive stress wave forms, causing cone cracking and radial/circumferential cracks that spread through the ceramic. The projectile is blunted or shattered, but the ceramic itself undergoes brittle fracture or localized amorphization (in B_4C). Backing materials (e.g., UHMWPE, Kevlar) are essential to capture fragments and absorb residual energy. Thus, the physical protection mechanism is a combination of projectile defeat and controlled ceramic failure, designed to prevent penetration while sacrificing the ceramic layer.

The ballistic efficiency of ceramics is strongly influenced by processing and microstructural control. Hot pressing and liquid-phase sintering of SiC improve grain boundary cohesion, increasing flexural strength and delaying brittle fracture. In alumina, higher purity and finer grain size reduce porosity, enhancing hardness and projectile erosion capability. Boron carbide requires sintering above 2000 °C, where microstructural defects or amorphization zones can reduce multi-hit performance. Thus, sintering method, grain size distribution, and porosity are critical factors that determine whether ceramics achieve their theoretical mechanical properties in real ballistic tests.

4. CONCLUSIONS

In this paper, some of the most used ceramics as bulletproof materials were presented. Some of the key applications, as well as the mechanical properties, were discussed. Based on the literature data, some strengths and weaknesses of the selected ceramics were highlighted.

Refrences

[1] G.M. Reddy and Manish Roy, Materials and Coatings for Defence Application, Defence Science Journal, Vol. 73, No. 2, March 2023, pp. 119-120, DOI:10.14429/dsj.73.18849

[2] Advanced Materials for Defense II, Development, Analysis and Application, Conference proceedings, 2020. DOI: 10.1007/978-3-030-34123-7, Publisher: Springer International Publishing, ISBN: 978-3-030-34122-0

[3] Harshit Mittal, Advanced Materials and Composites Revolutionizing Bullet-Resistant Vests, Brief Report, Preprints.org (www.preprints.org) Posted: 16 July 2025, doi:10.20944/preprints202507.1313.v1

[4] https://www.usbulletproofing.com/bulletproof-vs-bullet-resistant-difference/ 9th April 2026.

[5] https://www.defenselite.com/blog/bulletproof-materials-ranked/9th April 2026.

[6] https://ggsceramic.com/news-item/the-6-best-bulletproof-ceramic-materials-performance-comparison-and-selection-analysis-in-2025,9th April 2026.

[7] M. V. Silva, D. Stainer, H. A. Al-Qureshi, O. R. K. Montedo3 and D. Hotza, Alumina-Based Ceramics for Armor Application: Mechanical Characterization and Ballistic Testing, Hindawi Publishing Corporation, Journal of Ceramics, Volume 2014, Article ID 618154, 6 pages http://dx.doi.org/10.1155/2014/618154

[8] R. E. Tressler, "An assessment of low cost manufacturing technology, for advanced structural ceramics and its impact on ceramic armor," Ceramic Transactions, vol. 134, pp. 451–462, 2002.

[9] R. Stevens and P. A. Evans, "Transformation toughening by polycrystalline zirconia," Transactions and Journal of the British Ceramic Society, vol. 83, no. 1, pp. 18–31, 1984.

[10] R. B. Heimann, Classic and Advanced Ceramics: From Fundamentals to Applications,Wiley-VCH, New York, NY, USA, 2010.

[11] M.S. Boldin, N.N. Berendeev, N.V. Melekhin, A.A. Popov, A.V. Nokhrin, V.N. Chuvildeev, Review of ballistic performance of alumina: Comparison of alumina with silicon carbide and boron carbide, Ceramics International Volume 47, Issue 18, 15 September 2021, Pages 25201-25213

[12] https://www.chemshun.com/Article/99-Alumina-Ceramic-The-Superior-Choice-for-Bulletproof-Armor-Materials_1.html

[13] Dauren Zhambakin, Madi Abilev, Almira Zhilkashinova, Aigerim Ichshanova and Leszek Łatka, Silicon Carbide Ceramics for Armor Applications: A Review of Sintering Methods and Additive Systems, Molecules 2026, 31, 1185, https://doi.org/10.3390/molecules31071185

[14] Vijay Muthuvelan, Inamul Hasan, Sivaramraj Murugan, Ganesh Radha Krishnan, Mani Kandan Arumugam, G. Ramanan, Beena Stanislaus Arputharaj, Subhav Singh, Vijayanandh Raja, Processing–property–performance relationships in ceramic, metallic, and polymer-fiber armor systems: A quantitative review of constitutive models, ballistic limits, and design frontiers, Next Materials, Volume 13, 2026, 102798, ISSN 2949-8228,

https://doi.org/10.1016/j.nxmate.2026.102798.

[15] Chen Z, Hou J, Gao F, Deng S, Xu Y, Qin J, Cui S, Kim JH. Analysis of bulletproof performance of structurally optimized ceramic composite armor through numerical simulation and live fire test. Sci Rep. 2024 Dec 30;14(1):31685. doi: 10.1038/s41598-024-80752-0. PMID: 39738139; PMCID: PMC11685454.

[16] Ma, Minghui & Yu, Yilei & Gao, Guangfa. (2026). Dynamic fracture damage mechanism and impact resistance of ceramics under impact load. Journal of Materials Science. 61. 4384-4400. 10.1007/s10853-025-12131-7.

[17] Sonia Aprilya1 and Sovian Aritonang, Comparison Of Al2O3, SiC, and B4C as Ballistic-Resistant Body Armor, International Journal of Progressive Sciences and Technologies (IJPSAT) ISSN: 2509-0119.Vol. 42 No. 2 January 2024 ISSN: 2509-0119

[18] Fangfang Qi, Cheng Wang, Wenlong Xu, Ballistic protection and damage mechanism of ceramic composite armor under two-dimensional pre-stressed constraints by molten metal casting, Aerospace Science and Technology, Volume 159, 2025, 110021, ISSN 1270-9638, https://doi.org/10.1016/j.ast.2025.110021.

[19] https://www.preciseceramic.com/blog/boron-carbide-key-properties-applications.html, 10.04.2026 4:35

[20] S.G. Savio, K. Ramanjaneyulu, V. Madhu, T. Balakrishna Bhat, An experimental study on ballistic performance of boron carbide tiles, International Journal of Impact Engineering Volume 38, Issue 7, July 2011, Pages 535-541

[21] Pedro Henrique Poubel Mendonça da Silveira, ,Thuane Teixeira da Silva, , Matheus Pereira Ribeiro, , Paulo Roberto Rodrigues de Jesus, , Pedro Craveiro Rodrigues dos Santos Credmann, , Alaelson Vieira Gomes, A Brief Review of Alumina, Silicon Carbide and Boron Carbide Ceramic Materials for Ballistic Applications, October 2021, Academia Letters, DOI: 10.20935/AL3742

[22] Sonia Aprilya and Sovian Aritonang, Comparison Of Al2O3, SiC, and B4C as Ballistic-Resistant Body Armor, International Journal of Progressive Sciences and Technologies (IJPSAT) ISSN: 2509-0119, Vol. 42 No. 2 January 2024, pp. 413-418

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ADSORPTION OF AN ORGANIC POLLUTANT ONTO BIOCHAR DERIVED FROM SOUR CHERRY STONE

ANJA ANTANASKOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia, a.antanaskovic@itnms.ac.rs

MILAN MILIVOJEVIĆ
University of Belgrade, Faculty of Technology and metallurgy, Belgrade, Serbia, mmilan@tmf.bg.ac.rs

NIKOLA VUKOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia, n.vukovic@itnms.ac.rs

VLADIMIR ADAMOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia, v.adamovic@itnms.ac.rs

TATJANA ŠOŠTARIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia, t.sostaric@itnms.ac.rs

ZORICA LOPIČIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, Serbia, z.lopicic@itnms.ac.rs

Abstract: *In the last decade, the expansion of industrial and agricultural activities has led to the significant rise of organic pollutants, contributing to serious environmental contamination. Biochar has received considerable attention as a low-cost and environmentally friendly adsorbent for a wide range of pollutants. In this paper, waste sour cherry stones, originating from the local fruit industry, were used as a raw material for biochar production via slow pyrolysis. The obtained biochar was characterized by SEM-EDX analysis, which revealed a porous and irregular surface morphology. The adsorption of the cationic dye Malachite green (MG) onto biochar was investigated in stirred batch system at pH 6, 25 °C, adsorbent dosage of 2 g/L, initial dye concentration of 15-400 mg/L for 24 h. To assess the adsorption capacity of biochar for MG, the experimental data were analyzed by non-linear fitting of the Langmuir, Freundlich, and Sips adsorption isotherm models. The maximum adsorption capacity of 116.33 mg/g was obtained by Sips model, indicating combined adsorption on heterogeneous surfaces. The utilization of waste sour cherry stones for biochar production offers an effective approach for reducing waste disposal problems and producing a novel and sustainable adsorbent with promising potential for removal of organic pollutants from aqueous solutions.*

Keywords: *lignocellulosic waste, pyrolysis, dye removal, isotherms*

1. INTRODUCTION

Global population growth, urbanization, industrial expansion, and evolving lifestyles have contributed to the growing release of potentially polluting substances into the environment, placing considerable strain on water resources. Worldwide assessments indicate that water pollution is a critical environmental challenge, with repercussions felt globally, posing severe threats to human health, food security, and ecosystem functioning [1]. Of particular concern are organic pollutants, whose persistent nature allows them to remain in water systems for extended periods. These pollutants include a wide range of compounds such as pharmaceuticals, dyes, pesticides, and industrial chemicals, as well as hazardous organic substances that may originate from the chemical industry and military activities. Due to their toxicity, persistence, semi-volatile nature, and resistance to natural biodegradation, these compounds pose significant risks to both ecosystems and human health [2,3]. Furthermore, the persistence of organic pollutants can lead to their bioaccumulation and biomagnification through food chains [4].

Among these contaminants, organic dyes constitute a significant class of pollutants due to their widespread industrial use and direct release into aquatic systems. In particular, Malachite Green (MG) has been extensively studied as a model dye, owing to its high toxicity, chemical stability, complex structure and resistance to conventional degradation processes. This dye is widely used in the textile industry, as a therapeutic agent in aquaculture, and for food-related applications [5]. Prolonged exposure to this dye has been associated with potential carcinogenic and mutagenic effects [6]. In order to effectively remove MG from contaminated water, various treatment technologies have been developed, among which adsorption has been widely recognized as one of the most effective methods due to its cost-effectiveness, simplicity, and operational flexibility [2,6].

In recent decades, various materials have been investigated as adsorbents for organic pollutants, while

Corresponding author: A. Antanasković;
E-mail: a.antanaskovic@itnms.ac.rs
DOI: 10.5937/OTEH26125A

biochar has emerged as a promising material for environmental applications. Biochar is a stable carbon-rich material produced through the pyrolysis of biomass under oxygen-limited conditions, which is widely recognized for its low cost, abundant feedstock availability, and versatile applications in environmental remediation [7,8]. Due to the highly porous structure, large specific surface area, and diverse surface functional groups (hydroxyl, phenolic and carboxyl groups), biochar exhibits excellent adsorption capacity toward various organic and inorganic contaminants [5,8].

Despite these advances, there is a growing need to develop feedstock-specific biochars that align with the principles of green chemistry and circular economy, transforming local agricultural by-products into high-value environmental solutions. In the present study, biochar produced from fruit waste industry was evaluated as a low-cost and sustainable adsorbent for the removal of MG from aqueous solutions and represents a basis for further investigations aimed at developing cost-effective and efficient adsorbents for organic pollutant removal. The results provide a basis for further research on the application of the obtained biochar for the removal of other organic pollutants, such as fuel residues and pesticides, from water contaminated by military activities.

2. MATERIALS AND METHODS

Sour cherry stones (*Prunus cerasus L.*) were obtained from the local fruit processing company "Fruvita", where they are generated as a waste by-product. The collected stones were separated from fruit residues, thoroughly washed with tap water, and dried at room temperature prior to further processing. The dried material was subsequently pyrolysed to produce biochar. Pyrolysis was carried out in a Nabertherm 1300 muffle furnace (Nabertherm, Germany) at 500 °C for 90 minutes under an argon atmosphere. After cooling, the produced biochar was ground and sieved to obtain a particle size fraction between 0.1 and 0.5 mm.

The yield of produced biochar is calculated according to the following Equation (1):

$$Yield(\%) = \frac{mass\ of\ pyrolised\ sampel\ (g)}{mass\ of\ raw\ sampel\ (g)}\ x\ 100 \qquad (1)$$

Produced biochar was characterised by using the SEM-EDX analysis by JEOL JSM-6610 LV SEM (JEOL Ltd., Japan). Prior to analysis, the samples were coated with a thin layer of gold. EDX measurements were conducted at an accelerating voltage of 20 kV, a probe current of approximately 13 nA, a working distance of 10 mm, a magnification of 500×, and an acquisition live time of 100 s. The elemental composition obtained by EDX analysis was expressed as weight percentages (wt. %).

Malachite green, also known as Basic Green 4 was supplied by Sigma-Aldrich, (USA) and used as received without further purification. A stock solution (1000 mg/L) was prepared by accurately dissolving the required amount of dye in distilled water, and working solutions of different concentrations were subsequently obtained by serial dilution of this stock solution.

The adsorption isotherms were performed in batch mode at 25 °C. Briefly, 0.1 g of biochar was added to 50 mL of MG solutions with initial concentrations ranging from 15 to 400 mg/L. The systems were equilibrated for 24 h, after which the residual dye concentration was determined spectrophotometrically by measuring the absorbance by UV/VIS spectrophotometer (Spekol 1300, Analytik Jena, Germany) at 614 nm. The adsorption capacity q (g/mg) was calculated using Equation (2):

$$q_e = \frac{V\ x\ (C_i - C_f)}{\mathrm{m}} \qquad (2)$$

where C_i and C_f are the equilibrium and initial concentrations of dye (mg/L), respectively, V is the volume of dye solution (L) and m is the amount of the sorbent (g).

The equilibrium data were analyzed using Langmuir, Freundlich, and Sips isotherm models, which are described in detail in our previous study [7]. The best fitting model was determined by the highest correlation coefficients (R^2) and the lowest chi-squared value (χ^2).

3. RESULTS AND DISCUSSION

Schematic illustration of the adsorbent preparation and adsorption experiment is provided in Figure 1.

Figure 1. Schematic diagram of the experiment

The obtained biochar yield was 26.73%, which is typical for biochar produced from fruit stones [9,10].

To investigate the surface morphology and elemental composition of the biochar, SEM–EDX analysis was performed and presented in Figure 2. The biochar exhibited irregular and heterogeneous morphology with visible pores of different sizes and shapes. The observed porous features may be associated with the thermal decomposition of lignocellulosic components, including cellulose and hemicellulose, during pyrolysis. The decomposition of these components is accompanied by the release of volatile compounds, which may contribute to the formation of voids and channels within the carbon matrix [11]. The EDX analysis indicated that carbon was the dominant element in the biochar (81.37%), followed by oxygen (18.44%). These results are in agreement with the study by Lopičić et al., [12] who investigated the SEM–EDX characteristics of biochar derived from plum stones.

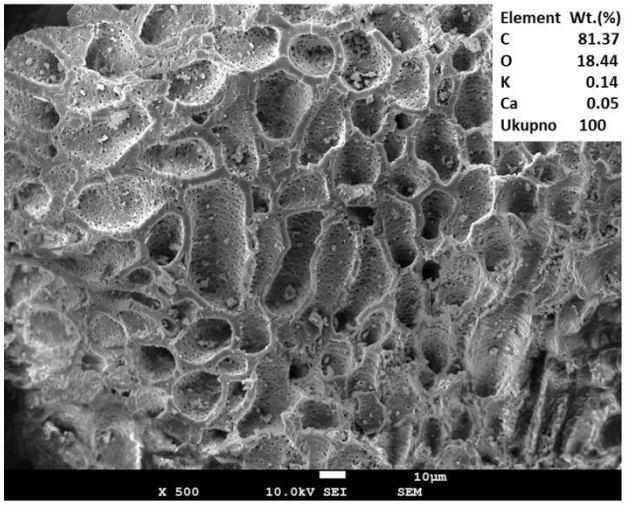


Figure 2. SEM-EDX micrographs of the biochar from Sour cherry stone

To elucidate the nature of the interaction between MG and biochar, adsorption isotherm analysis was performed. The experimental data were fitted using the Langmuir, Freundlich, and Sips isotherm models and presented in Figure 3, while the corresponding parameters are listed in Table 1.

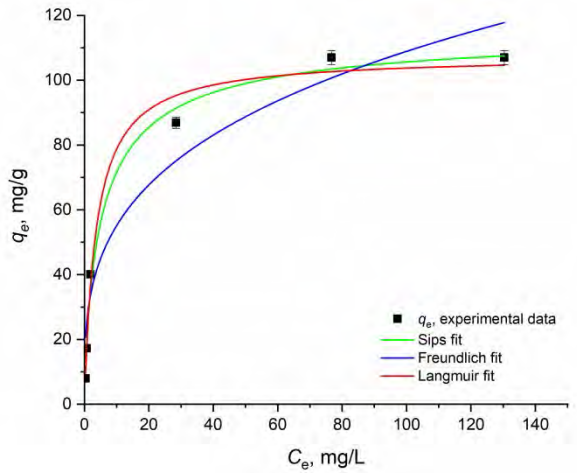


Figure 3. Isotherm model fitting of MG adsorption on biochar

The suitability of the applied isotherm models for describing MG adsorption was evaluated using the correlation coefficient (R^2) and chi-square (χ^2) values as statistical criteria. According to the data presented in Table 1, the Sips model provided the best fit for MG adsorption onto biochar. The Sips model exhibited the highest R^2 value (0.991) and the lowest χ^2 value (2.657), indicating that MG adsorption onto biochar occurs on heterogeneous surface sites with varying energies. The Sips heterogeneity parameter s was 0.791, which confirms a heterogeneous adsorption surface with non-uniform binding energies. The maximum predicted adsorption capacity obtained from the Sips model was 116.332 mg/g. Das et al., [5] reported the similar adsorption capacity of 120.50 mg/g for MG removal by NaOH-activated biochar derived from agricultural waste. Considering that the biochar used in this study was not activated, the obtained results highlighting the potential of the biochar derived from waste sour cherry stones for further optimization and application in the removal of organic pollutants from aqueous solutions. Future studies should evaluate the applicability of the biochar in real water matrices, where competing dissolved components may affect the adsorption process. The desorption of the adsorbed dye, together with the regeneration and reuse of the biochar, should also be evaluated to determine the potential for repeated application and the economic feasibility of the adsorption process. In addition, the obtained biochar may be further investigated as a low-cost adsorbent for the treatment of contaminated water resulting from military activities, particularly where locally available waste-derived materials may be advantageous.

Table 1. Isotherm parameters for the adsorption of MG onto biochar

Isotherm model	Parameter	Value
Langmuir	q_e, mg/g	107.551
	K_L	0.273
	R^2	0.985
	$\chi 2$	7.374
Freundlich	K_F	27.904
	n	0.296
	R^2	0.931
	χ^2	9.696
Sips	q_e, mg/g	116.33
	K_S	0.259
	s	0.791
	R^2	0.991
	χ^2	2.657

4. CONCLUSION

This study demonstrates that biochar derived from agro-industrial waste can be used as an efficient, low-cost, and sustainable adsorbent for the removal of organic pollutants from aqueous systems, such as MG. Conversion of fruit industry waste into a functional adsorbent not only provides an effective approach for wastewater treatment, but also contributes to waste valorization and the reduction of environmental burdens in accordance with the principles of the circular economy.

Among the investigated isotherm models, the Sips model showed the best agreement with the experimental data, indicating heterogeneous adsorption behavior, with a maximum adsorption capacity of 116.33 mg/g. The high adsorption capacity may be associated with the porous morphology observed by SEM, which may facilitate the uptake of relatively large dye molecules. These findings provide a basis for future studies focused on the optimization, potential surface modification, and application of biochar derived from sour cherry stones for the removal of organic pollutants from aqueous solutions.

Acknowledgement

This work is supported by the Ministry of Education, Science and Technology Development of the Republic of Serbia (contract number: 451-03-33/2026-03/200023).

References

[1] MUKARUNYANA, B.: *Biochars and Hydrochars for the Adsorption of Organic Contaminants from Wastewater*, Umeå University, Sweden, 2026.

[2] BHOMICK, P.C., SUPONG, A., SINHA, D.: *Organic Pollutants in Water and Its Remediation Using Biowaste Activated Carbon as Greener Adsorbent,* International Journal of Hydrology, 1(3) (2017) 91-92.

[3] JAGADEESH, N., SUNDARAM, B.: *Adsorption of Pollutants from Wastewater by Biochar: A Review*, Journal of Hazardous Materials Advances, 9(2023) 100334.

[4] BELJIN, J., ĐUKANOVIĆ, N., ANOJIĆ, J., APOSTOLOVIĆ, T.: Biochar in the Remediation of Organic Pollutants in Water: A Review of Polycyclic Aromatic Hydrocarbon and Pesticide Removal, Nanomaterials, 15(1) (2025) 26.

[5] DAS, T., DEBNATH, A., SEKHAR, M.: *Adsorption of Malachite Green by Aegle Marmelos-Derived Activated Biochar: Novelty Assessment through Phytotoxicity Tests and Economic Analysis*, Journal of the Indian Chemical Society, 101(9) (2024) 101219.

[6] PARK, J., WANG, J.J., MENG, Y., WEI, Z., DELAUNE, R.D., SEO, D.: *Adsorption/Desorption Behavior of Cationic and Anionic Dyes by Biochars Prepared at Normal and High Pyrolysis Temperatures*, Colloids and Surfaces A, 572 (2019) 274–282.

[7] ANTANASKOVIĆ, A., LOPIČIĆ, Z., PEHLIVAN, E., ADAMOVIĆ, V., ŠOŠTARIĆ, T., MILOJKOVIĆ, J., MILIVOJEVIĆ, M.: *Thermochemical Conversion of Non-Edible Fruit Waste for Dye Removal from Wastewater,* Biomass Conversion and Biorefinery, 14 (2024) 18649–18665.

[8] DONG, B., HUAN, Z., CAI, L., LIU, L., HAN, M., NIE, G., ZHAO, S., LIU, G., ZHU, Y.: *Biochar Applications for Efficient Removal of Energetic Compound Contaminants*, Chemosphere, 364 (2024) 143135.

[9] POLLARD, Z.A., GOLDFARB, J.L.: *Valorization of Cherry Pits: Great Lakes Agro-Industrial Waste to Mediate Great Lakes Water Quality*, Environmental Pollution, 270 (2021) 116073.

[10] LOPIČIĆ, Z., ANTANASKOVIĆ, A., ŠOŠTARIĆ, T., ADAMOVIĆ, V., ORLIĆ, M.: Improvement of Energy Properties of Lignocellulosic Waste by Thermochemical Conversion into Biochar, Hemijska industrija, 77(2) (2025) 147–154.

[11] BANCZEK, P., BRUGNERA, A.B., CHRISTOFORO, A.L., MACHADO, G.D.O.: *Morphological and Elementary Evaluation of Wooden Carbonaceous Materials from Activated Carbon Industry,* Nativa, Sinop, 7(2) (2019) 213–217.

[12] LOPIČIĆ, Z.R., ŠOŠTARIĆ, T.D., MILOJKOVIĆ, J.V., ANTANASKOVIĆ, A.V., MILIĆ, J.S., SPASIĆ, S.D., AVDALOVIĆ, J.S.: *Efficient Removal of Water Soluble Fraction of Diesel Oil by Biochar Sorption Supported by Microbiological Degradation*, Processes, 12 (2024) 1962.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF PARTICLE SIZE ON ADSORPTION PROPERTIES OF GROUND FOAM GLASS

VELJKO SAVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, v.savic@itnms.ac.rs

ANJA ANTANASKOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, a.antanaskovic@itnms.ac.rs

TAMARA VUJATOVIĆ-VELIMIROV
Military Technical Institute, Belgrade, tamara.vujatovic@gmail.com

VLADIMIR TOPALOVIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, v.topalovic@itnms.ac.rs

ZORICA LOPIČIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, z.lopicic@itnms.ac.rs

JELENA NIKOLIĆ
Institute for Technology of Nuclear and Other Mineral Raw Materials, Belgrade, j.nikolic@itnms.ac.rs

SNEŽANA GRUJIĆ
University of Belgrade, Faculty of Technology and Metallurgy, Belgrade, grujic@tmf.bg.ac.rs

Abstract: *Foam glass is an inorganic insulation material composed of a glassy matrix with gas-filled pores. In this study, it was produced entirely from secondary raw materials: green container glass as the glassy matrix and sugar beet factory lime as the foaming agent. During sintering and foaming, the material expands irregularly, and it should be cut into blocks. Ground and sieved foam glass can be used as an adsorbent. This study investigates the effect of particle size on the adsorption capacity of Brilliant Green (BG). Four particle size fractions were examined (<0.2 mm, 0.2-0.5 mm, 0.5-0.8 mm, and >0.8 mm) and characterized by BET analysis to evaluate the variation of specific surface area with particle size. The adsorption experiments were performed in a batch system at room temperature for 24 h, using an adsorbent dosage of 2 g/L and a BG concentration of 50 mg/L. The smallest particle size fraction (<0.2 mm) showed the highest dye removal efficiency (96%), attributed to its largest specific surface area, while further increases in particle size led to a decrease in removal efficiency. Ground foam glass waste was successfully repurposed as an adsorbent, promoting circular economy principles.*

Keywords: *Foam glass, Adsorption, Brilliant Green, Recycling*

1. INTRODUCTION

The rapid expansion of industrial activities has led to the increased discharge of synthetic dyes into water systems, posing significant environmental and public health concerns. Dyes are used in textile, paper, leather, printing, pharmaceutical, and cosmetic industries because of their excellent stability, high coloring efficiency, and resistance to degradation. However, these same properties make many dyes persistent pollutants that are difficult to remove by conventional wastewater treatment processes. Even at low concentrations, colored effluents reduce light penetration in water bodies, inhibit photosynthetic activity, disrupt aquatic ecosystems, and negatively affect water quality [1].

Among synthetic dyes, Brilliant Green (BG) has attracted considerable attention due to its widespread use and high toxicity. Brilliant Green is a cationic triphenylmethane dye. It has been widely used as a biological stain, veterinary antiseptic, disinfectant, and colorant in the textile and paper industries. Despite its practical applications, the discharge of BG into water systems represents a serious environmental challenge because of its high chemical stability and poor biodegradability. The dye can persist in water for extended periods, increasing the risk of accumulation in natural ecosystems [2].

The presence of BG in wastewater is associated with numerous adverse effects on living organisms. Exposure to this dye has been reported to cause skin and eye irritation, respiratory disorders, gastrointestinal disturbances, and allergic reactions. Prolonged exposure has been linked to mutagenic, cytotoxic, and potentially carcinogenic effects. In aquatic environments, BG can

Corresponding author: V. Savić;
E-mail: v.savic@itnms.ac.rs
DOI: 10.5937/OTEH26126S

inhibit the growth of algae and microorganisms, disrupt food chains, and reduce dissolved oxygen levels by limiting light penetration and photosynthetic activity. Consequently, the efficient removal of BG from contaminated water has become an important objective in environmental protection and wastewater management [2].

Among the available treatment technologies, adsorption has emerged as one of the most promising methods for dye removal because of its simplicity, high efficiency, ease of operation, and economic feasibility. The adsorption process involves the accumulation of dye molecules on the surface of a solid adsorbent through physical or chemical interactions, allowing rapid and efficient removal even at low pollutant concentrations. Compared with other treatment methods, adsorption offers several advantages, including high removal efficiency, low energy requirements, minimal formation of harmful by-products, operational flexibility, and the possibility of adsorbent regeneration and reuse [1,2].

In recent years, considerable research efforts have focused on the development of low-cost, sustainable, and high-performance adsorbents derived from natural minerals, industrial by-products, and agricultural residues [1]. These materials often exhibit large specific surface areas, abundant active adsorption sites, and favorable surface chemistries. The utilization of waste-derived adsorbents is attractive because it simultaneously addresses wastewater purification and waste valorization, contributing to the principles of circular economy and sustainable resource management.

For dye removal from aqueous solution many types of adsorbents were examined (clays, biochars, etc.), but there is limited research on glass or foam glass application. Foam glass is a highly porous inorganic material consisting of a continuous glassy matrix enclosing a network of gas-filled pores. It is typically produced from finely ground waste glass cullet mixed with a foaming agent, most commonly calcium carbonate ($CaCO_3$), carbon (C) or silicon carbide (SiC). During heat treatment, the glass particles undergo viscous flow sintering while the foaming agent decomposes or reacts to generate gases. The resulting increase in internal gas pressure promotes pore nucleation, growth, and, in some cases, pore coalescence, leading to the expansion of the material [3].

Foam glass exhibits low density, good mechanical strength and excellent thermal insulation properties. In addition, its excellent chemical durability minimizes the release of hazardous substances into the environment during service, contributing to its long-term environmental stability [3].

Application of ground foam glass provides a low-cost and sustainable adsorbent for water purification, thereby integrating waste valorization with environmental remediation. The potential application of foam glass extends beyond dye removal, as its porous structure and surface properties make it a promising adsorbent for the removal of various pollutants from contaminated water. The possible use of ground foam glass may also extend to water treatment following floods or other events associated with water contamination. The aim of the present study is to investigate the feasibility of utilizing ground foam glass as an adsorbent for the removal of BG from aqueous solutions.

2. MATERIALS AND METHODS

Green container glass cullet was initially crushed in an agate mortar and subsequently milled in a Tencan planetary ball mill at 400 rpm for 60 min. Sugar beet factory lime (SBFL), a calcium carbonate-rich industrial by-product, was also crushed in an agate mortar and used as the foaming agent. The glass powder and SBFL were sieved using a No. 300 sieve (48 µm), and only particles smaller than 48 µm were used for foam glass production [3].

The glass powder was mixed with SBFL to obtain a batch containing 2.5wt.% $CaCO_3$. The powder mixture was uniaxially pressed into cylindrical pellets and sintered in a Carbolite CWF 13/13 electric furnace at 800 °C for 30 min using a heating rate of 10 °C/min [3]. After cooling to room temperature, the produced foam glass was crushed and sieved into four particle size fractions: d_1 (<0.2 mm), d_2 (0.2-0.5 mm), d_3 (0.5-0.8 mm), and d_4 (>0.8 mm). These fractions were subsequently used as adsorbents.

The surface area of the samples was analysed using N_2 physisorption at 77 K with an Anton Paar Nova 600. Before the analysis, samples were vacuum-degassed for 1 hour at room temperature, after which the temperature was increased to 100 °C and degassing continued for an additional 2 hour; then the temperature was increased again to 300 °C and degassing continued for 4 hours. When the temperature dropped to room temperature, vacuum degasation continued for additional 2 hours. The specific surface area of the samples was calculated from the linear portion of the adsorption isotherm branch by applying the Brunauer-Emmett-Teller (BET) equation [4-6]. The specific surface area for sample d_4 (>0.8 mm) could not be determined, likely due to its low specific surface area.

Batch adsorption experiments were conducted in 250 mL Erlenmeyer flasks. For each experiment, 0.10 g of foam glass of the selected particle size fraction was mixed with 50 mL of a 50 mg/L BG solution and agitated for 24 h. The pH of dye solution was 6.25. The residual BG concentration in the solution was determined using a UV–Vis spectrophotometer (Spekol 1300, Analytik Jena, Germany) at a wavelength of 625 nm. All experiments were done in doublets and average value is presented. The solution for UV-Vis measurements was sampled after the adsorbent had settled. The adsorption performance was evaluated in terms of the dye removal efficiency (R, %) were calculated using Eq. (1):

$$R(\%) = \frac{C_i - C_f}{C_i} \cdot 100\% \quad (1)$$

where C_i and C_f are the initial and final concentrations of Brilliant Green in the solution (mg/L).

3. RESULTS AND DISCUSSION

Figure 1 shows the SEM micrograph of the foam glass produced using sugar beet factory lime as the foaming agent. The material exhibits a highly porous cellular structure consisting nearly spherical pores embedded within a continuous glassy matrix.

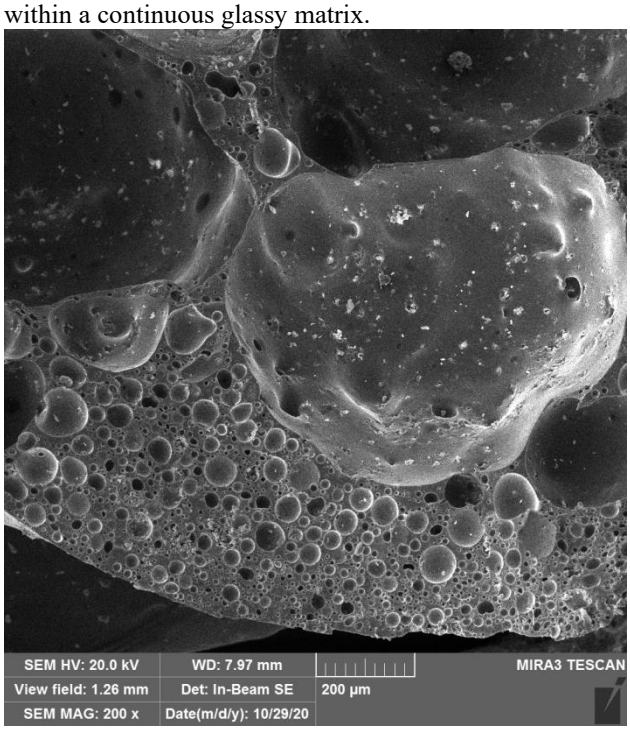


Figure 1. SEM micrograph of foamed glass sample

The pore structure is characterized by a broad pore size distribution. Numerous small pores are present throughout the structure, while larger pores are also observed, indicating that bubble nucleation and growth occurred simultaneously during foaming. The coexistence of pores of different sizes suggests that, in addition to the formation of new gas bubbles, some bubbles continued to grow. Several large pores appear to result from the coalescence of neighboring bubbles following the rupture of thin glass walls [3]. Despite the presence of several large pores, the majority of the cellular structure is composed of fine pores distributed relatively uniformly throughout the sample. The rough internal surfaces and numerous pore openings exposed after crushing are expected to provide accessible adsorption sites and facilitate the diffusion of dye molecules into the porous network.

The specific surface area was calculated from the data of absorption measurements by applying the BET equation (S_{BET}), and it is presented in Table 1.

Table 1. Specific surface area of ground foam glass samples

Sample	S_{BET} (m^2/g)	Correlation coefficient, R^2
d_1	0.194	0.9998
d_2	0.084	0.9999
d_3	0.055	0.9995
d_4	N.D.	N.D.

The BET specific surface area of the foam glass samples ranged from 0.055 to 0.194 m²/g, indicating relatively low nitrogen-accessible surface areas. An increase in particle size resulted in a decrease in the specific surface area. The observed increase in BET surface area with decreasing particle size is attributed to particle fragmentation during crushing, which exposes fresh fracture surfaces and opens previously inaccessible pores. Smaller particles therefore provide a greater external surface area per unit mass and a larger number of accessible adsorption sites. Such values are characteristic of soda-lime glass materials and are consistent with those reported for both waste glass and foam glass in the literature. Papelis et al. [7] reported BET surface areas between 0.0016 and 1.65 m²/g for various natural and manmade glasses, Similarly, Katsuura et al. [8] measured a BET surface area of only 0.04 m²/g for foam glass despite the presence of numerous pores with diameters between 10 and 900 µm. The adsorption performance of the ground foam glass is expected to depend not only on the measured BET surface area but also on factors such as particle size, pore accessibility, surface chemistry, and electrostatic interactions between the glass surface and BG molecules [1,2].

Figure 3 shows removal efficiency of different particle size of ground foam glass of BG.

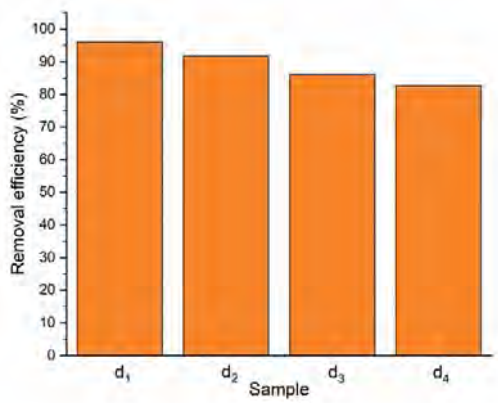


Figure 3. Removal efficiency of BG dye

The highest removal efficiency (96%) was achieved by the smallest particle size fraction (<0.2 mm), whereas progressively lower removal efficiencies were observed for larger particles.

Removal efficiency increase is accompanied by an increase in BET specific surface area from 0.055 to 0.194 m²/g with decreasing particle size, which could indicate that particle comminution increases the surface available for adsorption.

The better removal efficiency of the finest particles can be possibly attributed to their larger external surface area [2]. Crushing the foam glass exposes fresh fracture surfaces and opens previously closed cells, thereby increasing the number of accessible adsorption sites. Although the BET surface area increased with decreasing particle size, the measured values remained relatively low compared with conventional porous adsorbents such as activated carbon or biochar [2].

In addition to the physical characteristics of the adsorbent, electrostatic interactions may likely play an important role in the adsorption mechanism. BG is a cationic dye, whereas the surface of soda-lime glass in aqueous solution is generally negatively charged due to the deprotonation of surface silanol (Si–OH) groups. Attractive electrostatic interactions between the positively charged dye molecules and negatively charged glass surface could possibly be reason for enhanced dye uptake.

5. CONCLUSIONS

Ground foam glass generated was successfully used as a low-cost adsorbent for the removal of BG from aqueous solutions. The foam glass produced from recycled green container glass and sugar beet factory lime exhibited a highly porous cellular structure composed predominantly of macropores formed during the foaming process. The produced grounded foam glass proved to be effective for dye removal, achieving removal efficiencies of 82-96%. The highest removal efficiency was obtained using sample d_1, which also exhibited the largest measured BET surface area. The higher adsorption performance of smaller particles is attributed to the greater accessible surface area, exposure of fresh fracture surfaces during crushing. This study presents preliminary research in which high removal efficiency of dye was achieved, providing a solid foundation for future research. By utilization of ground foam glass material for wastewater treatment, this work contributes to both sustainable waste management and the development of environmentally friendly adsorption technologies.

Acknowledgments

This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-33/2026-03/200023, 451-03-34/2026-03/200135, 451-03-34/2026-03/200325, and 451-03-33/2026-03/200287).

References

[1] AL-KROOM, H., ABDEL-GAWWAD, H.A., ELRAHMAN, M.A., ABDEL-ALEEM, S., AHMED, M.S., SALAMA, Y.F., QAYSI, S., TECHMAN, M., SELIEM, M.K., YOUSSF, O.: Utilization of Foamed Glass as an Effective Adsorbent for Methylene Blue: Insights into Physicochemical Properties and Theoretical Treatment, Materials, 16 (2023), 1412.

[2] ANTANASKOVIĆ, A., LOPIČIĆ, Z., PEHLIVAN, E., ADAMOVIĆ, V., ŠOŠTARIĆ, T., MILOJKOVIĆ, J., MILIVOJEVIĆ, M.: Thermochemical Conversion of Non-Edible Fruit Waste for Dye Removal from Wastewater, Biomass Conversion and Biorefinery, 14 (2024), 18649–18665.

[3] SAVIĆ, V., TOPALOVIĆ, V., NIKOLIĆ, J., JEVTIĆ, S., MANIĆ, N., KOMATINA, M., MATIJAŠEVIĆ, S., GRUJIĆ, S.: Foam Glasses Made from Green Bottle Glass and Sugar Beet Factory Lime as a Foaming Agent, Heliyon, 9 (2023), e17664.

[4] ROUQUEROL, J., AVNIR, D., FAIRBRIDGE, C.W., EVERETT, D.H., HAYNES, J.H., PERNICONE, N., RAMSAY, J.D.F., SING, K.S.W., UNGER, K.K. Recommendations for the characterization of porous solids [IUPAC Recommendations 1994]. Pure Appl. Chem. 1994, 66, pp. 1739-1758.

[5] BRUNAUER, S., EMMETT, P.H., TELLER, E. Adsorption of gases in multimolecular layers. J. Am. Chem. Soc. 1938, 60, pp. 309-319.

[6] ISO 9277:2010, Determination of the specific surface area of solids by gas adsorption — BET method, 2nd edition, 210-09-01

[7] PAPELIS, C., UM, W., RUSSELL, C.E., CHAPMAN, J.B.: Measuring the Specific Surface Area of Natural and Manmade Glasses: Effects of Formation Process, Morphology, and Particle Size, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 215 (2003), 221–239.

[8] KATSUURA, Y., ASAOKA, S., TAKEDA, K., NAKASHITA, S., HAYASHI, K., TANAKA, K., INADA, Y., OKUDA, T.: Adsorptive Removal of Phosphate from Aqueous Solutions Using Iron–Lanthanum-Doped Foam Glass Adsorbent, Water Environment Research, 97 (2025), 1–10

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

RADIOACTIVITY IN FOOD SAMPLES MEASURED IN THE RADIOCHEMICAL LABORATORY OF THE MILITARY TECHNICAL INSTITUTE

NATAŠA PAJIĆ
Military Technical Institute, Belgrade, natasa.pajic969@gmail.com

ANTON RADOJKOVIĆ
Military Technical Institute, Belgrade, radojkovicanton@gmail.com

Abstract: *Environmental radioactivity monitoring includes the assessment of radiation exposure through food consumption, which represents a significant pathway for human intakes. The determination of radionuclide activity concentrations in foodstuffs provides essential data for evaluating background radiation levels and identifying potential contamination originating from anthropogenic sources, including nuclear and other radiation–related facilities .The Radiochemical Laboratory of the Military Technical Institute performs systematic gamma spectrometric measurements of representative food samples obtained from the Military Medical Academy, contributing to the system of radiological protection within the defence sector. This study presents the results of activity concentration measurements in 20 food samples, including dairy products, cereals fruits ,vegetables, eggs, and meat products used in standard daily diet. The analysis was performed using high-resolution gamma spectrometry. The aim of this study is to assess the natural radionuclides activity concentration in commonly consumed food and to evaluate the associated radiological health risk. An overview of monitoring results from 2015 to 2020 is also presented, providing insight into long-term trends and supporting compliance with relevant radiological safety standards.*

Keywords: *radioactivity, food, monitoring, gamma spectrometry, effective dose*

1. INTRODUCTION

The presence of anthropogenic radionuclides in the environment of Serbia is primarily a consequence of the nuclear accident that occurred Chernobyl nuclear power plant accident in 1986. Several decades after this event, the radionuclide ^{137}Cs continues to be detected in various environmental compartments owing to its relatively long half-life of approximately 30 years.

In addition to the consequences of nuclear accidents, numerous anthropogenic activities contribute to increased levels of environmental radioactivity. These include nuclear energy production, nuclear weapons testing, fossil fuel combustion, mining activities, as well as the production and application of phosphate fertilizers. These processes may promote the release and redistribution of naturally occurring radionuclides within the environment.

Among the naturally occurring radionuclides relevant for environmental and food radioactivity monitoring are ^{40}K, ^{232}Th, and ^{238}U. These radionuclides are introduced into ecosystems through the exploitation and utilization of mineral resources, most notably through agricultural practices.

The primary objective of radiological monitoring of the food chain is to determine the occurrence of natural and anthropogenic radionuclides in food products and their transfer to the production environment. Such investigations involve the analysis of soil, animal feed, agricultural crops, milk and dairy products, eggs, meat from domestic and wild animals, as well as fish, in order to assess the potential transfer of radionuclides through the food chain to the final consumers [1].

The presence of naturally occurring radionuclides in the Earth's crust directly influences the food chain. Agricultural crops cultivated on such soils can absorb these radionuclides, leading to their accumulation in food products. Consequently, the consumption of contaminated food may result in internal radiation exposure of the human population. Therefore, monitoring radioactivity levels in food represents a key component in the assessment of the overall radiation exposure associated with daily dietary intake.

2. RADIOACTIVITY IN FOOD PRODUCTS AND MEASUREMENT METHODS

The intake of radionuclides through food represents a significant pathway for long-term exposure to ionizing radiation, potentially affecting various organs. Therefore, assessing the level of radioactive contamination in food products is of great importance for evaluating food safety and the radiological risk to human health.

The high natural abundance of potassium (^{40}K) in soil constitutes the largest contributor to internal radiation exposure in humans. Its beta and gamma emissions

Corresponding author: N. Pajić;
E-mail: natasa.pajic969@gmail.com

DOI: 10.5937/OTEH26127P

directly increase the level of internal radiation dose following ingestion. In this study, gamma spectrometry was employed to quantify the activity concentrations of naturally occurring radionuclides in various food products used for the preparation of daily meals in the kitchen of the Military Medical Academy.

Sources of radioactivity in food products can be classified into four main categories:

- Naturally Occurring Sources of Radioactivity,
- Industrial and Energy-Related Sources of Radioactivity,
- Radioactivity as consequence of nuclear weapons testing and nuclear accidents.

2.1 Naturally Occurring Sources of Radioactivity

Certain radionuclides represent an integral part of the natural environment. Among them, potassium (^{40}K) is a naturally occurring radioactive isotope of an element essential for biological processes. Uranium, thorium, and their decay progenies are also present in soil, water and the atmosphere. Plants absorb these elements from their surroundings and therefore trace amounts can be detected in fruits, vegetables, cereals, milk and meat. Monitoring natural radiation enables the assessment of population exposure and the timely detection of potential adverse changes. [2]

2.2 Industrial and Energy-Related Sources of Radioactivity

During the operation of nuclear facilities, fuel processing activities, laboratory operations, as well as certain medical procedures, small quantities of radionuclides may be released into the environment. These radionuclides can subsequently be transferred to water, soil, and the food chain. Although, such emissions are generally strictly regulated and maintained at low levels, continuous monitoring remains essential for the protection of human health.

2.3 Radioactivity as a Consequence of Nuclear Weapons Testing and Nuclear Accidents

Nuclear weapons testing during the second half of the 20th century resulted in the global dispersion of radioactive particles. These particles were deposited on soil, vegetation, and in aquatic environments, subsequently entering the food chain and spreading through various components of the environment. Certain radionuclides can still be detected decades after the conducted tests. [2]

2.4 Radioactivity in Food Products

Continuous monitoring of radioactivity in food products enables the timely detection of changes in the environment, the assessment of population exposure, and the implementation of appropriate protective measures if radionuclide concentrations exceed the established reference values. The results of long-term studies indicate that radionuclide activity concentrations in food are generally low in most countries and remain within levels that do not represent a significant risk to public health. Nevertheless, regular monitoring remains an essential component of the food safety system, particularly in the event of nuclear accidents or other emergency situations that may result in environmental contamination.

2.5 The Most Relevant Radionuclides in Food Products

- Radionuclides determined in food products can be classified into naturally occurring and artificial radionuclides. Their importance in environmental monitoring depends not only on their presence in the environment but also on their half-life, pathways of entry into the food chain, chemical properties, and contribution to the total effective radiation dose received by the population.
- Potassium-40 (^{40}K) is the most abundant naturally occurring radionuclide in food. As a naturally occurring isotope of potassium, it is present in all plant- and animal-derived food products. Since potassium is an essential element for the normal functioning of the human body, its concentration in tissues is regulated by physiological mechanisms. Consequently, ^{40}K is a major contributor to natural internal radiation exposure.
- Radium-226 (^{226}Ra) belongs to the uranium-238 (^{238}U) decay series and may be taken up by plants from soil or enter drinking water from geological formations. Its presence is of particular concern in groundwater and in certain food products originating from areas with elevated concentrations of naturally occurring radionuclides.
- Uranium-238 (^{238}U) and thorium-232 (^{232}Th) occur naturally in soils at low concentrations. Their transfer to plants is generally limited; however, elevated activity concentrations may be observed in certain geological regions, which is why these radionuclides are included in environmental monitoring programs.
- Polonium-210 (^{210}Po) is of particular importance in marine ecosystems. This radionuclide can accumulate in fish and seafood and may constitute one of the major sources of natural internal radiation exposure for populations with a high consumption of marine food products. [3]

2.6 Artificial Radionuclides

The most significant artificial radionuclide in food products is caesium (^{137}Cs). Following its release into the environment, it binds strongly to soil particles; however, it can be taken up by plants through their root systems. Its transfer through the food chain depends on soil characteristics, plant species, and climatic conditions. Owing to this, its may remain in the environment for several decades following a nuclear accident.

Strontium-90 (^{90}Sr) enters the human body primarily through the consumption of milk and dairy products because it exhibits chemical behavior similar to that of calcium. After absorption, it is incorporated into bone tissue, making its monitoring particularly important in foods intended for children.

In addition to these radionuclides, **iodine (^{131}I)**, **tritium (^{3}H)**, **plutonium-239 (^{239}Pu)**, and **americium-241 (^{241}Am)** may also be determined under specific circumstances. Although ^{131}I has a relatively short half-life (approximately eight days), it may pose a significant health risk following nuclear accidents because of its accumulation in the thyroid gland. Consequently, milk, drinking water, and leafy vegetables are subject to intensive monitoring immediately after such events.

Table 1. The most relevant radionuclides in food

Radionuclide	Origin	Main Food Products	Monitoring Significance
^{40}K	Natural	Fruits, vegetables, cereals, milk, meat	natural internal radiation exposure
^{226}Ra	Natural	Drinking water, vegetables	assessment of regional geological characteristics
^{238}U	Natural	Drinking water, cereals	assessment of natural radioactivity
^{232}Th	Natural	Plant-based foods	natural background radioactivity
^{210}Po	Natural	Fish and seafood	dietary intake from aquatic ecosystems
^{137}Cs	Artificial	Milk, meat, mushrooms, game meat, cereals	primary indicator of nuclear contamination
^{90}Sr	Artificial	Milk and dairy products	accumulation in bone tissue
^{131}I	Artificial	Milk, leafy vegetables	post-nuclear accident monitoring

The selection of radionuclides included in routine monitoring programs depends on the national monitoring strategy, the contamination history of a particular region, and the assessment of potential risks to the population. Under normal conditions, the greatest emphasis is placed on determining the activity concentrations of ^{40}K and ^{137}Cs, whereas other radionuclides are analyzed as required or within the framework of specialized monitoring and research programs.

2.7 Food Radioactivity Monitoring

Food radioactivity monitoring is an organized and continuous system of sampling, laboratory analysis, and assessment of radionuclide activity concentrations in food products. Its primary objective is to protect public health through the timely detection of potential food contamination and the assessment of exposure to ionizing radiation via dietary intake. In addition to routine surveillance, monitoring enables a rapid response in the event of nuclear or radiological accidents when there is an increased likelihood of radionuclides entering the food chain.

Monitoring programs are based on the recommendations of the International Atomic Energy Agency (IAEA). In the Republic of Serbia, environmental and food radioactivity monitoring is conducted in accordance with the Law on Radiation and Nuclear Safety and Security, as well as the relevant other regulations that define the types of samples, the frequency of testing, and the applicable reference values.

The selection of food products included in monitoring programs is based on their importance in the daily diet of the population, their capacity to accumulate radionuclides, and their suitability as indicators of environmental changes. Particular attention is given to milk, drinking water, cereals, leafy vegetables, and infant foods, as these products may contribute significantly to the overall dietary intake of radionuclides among sensitive population groups.

2.8 Radionuclide Determination Methods

Laboratory analysis of food radioactivity is performed using standardized methods that enable the reliable determination of the activity concentrations of individual radionuclides.

The most commonly used methods include:

- High-resolution gamma spectrometry (HPGe detectors) – used for the determination of gamma-emitting radionuclides such as ^{40}K, ^{137}Cs, ^{60}Co, and ^{131}I. The main advantage of this method is that it allows the simultaneous determination of multiple radionuclides without chemical preparation of the sample.
- Radiometric determination of beta-emitting radionuclides – applied for the determination of ^{90}Sr following appropriate chemical separation procedures.
- Alpha spectrometry – used for the determination of alpha-emitting radionuclides such as ^{210}Po, ^{238}U, and ^{239}Pu.

The types of food products subject to monitoring vary between countries, however, the main categories are generally similar due to their importance in human nutrition and their potential for radionuclide accumulation.

Table 2. Major food groups included in food radioactivity monitoring programs

Food Group	The Most Frequently Monitored Radionuclides	Reason for Inclusion in Monitoring
Milk and dairy products	^{40}K, ^{137}Cs, ^{90} Sr, and ^{131}I	Highly sensitive indicator of contamination;

		important for infant and child nutrition
Cereals and cereal products	^{40}K, ^{137}Cs	Significant contribution to the daily diet
Bread and flour	^{40}K, ^{137}Cs	Control of essential staple foods
Leafy vegetables	^{137}Cs, ^{131}I *, ^{40}K	Direct deposition of radionuclides from the atmosphere
Fruits	^{40}K, ^{137}Cs	Assessment of intake through plant-based foods
Meat from domestic animals	^{40}K, ^{137}Cs	Transfer of radionuclides through animal feed
Fish and seafood	^{210}Po, ^{137}Cs	Monitoring of radioactivity in aquatic ecosystems
Mushrooms	^{137}Cs	High capacity for cesium accumulation
Honey	^{40}K, ^{137}Cs	Indicator of environmental conditions
Drinking water	^{226}Ra, ^{238}U, ^{3}H	Assessment of exposure through drinking water

The results of monitoring are used for the assessment of the annual effective dose received by the population, evaluation of food safety, and planning of protective measures in the event of increased radionuclide activity concentrations.

3. MATERIAL AND METHODS

Monitoring was performed on food product samples obtained through a joint initiative of the Military Technical Institute and the Military Medical Academy. The samples consisted of food items used to prepare daily meals for members of the Serbian Armed Forces, and were included in a comprehensive radiological, chemical, and biological monitoring program.

The study was conducted between 2015 and 2020. The analyzed samples represented a random selection of food products collected as part of the routine monitoring program. Sample preparation was performed at the Military Medical Academy laboratory using standard Marinelli geometry. Moist samples were treated with hydrochloric acid prior to analysis. Following this treatment, the containers were left open for 24 hours before sealing to allow the escape of radon gas. Subsequently, the samples were hermetically sealed with Parafilm®, appropriately labelled, and stored for a minimum of 28 days to establish secular equilibrium between radionuclides of ^{238}U and ^{232}Th and their short-lived decay products. After equilibration, the samples were subjected to gamma spectrometric analysis in the Radiochemistry Laboratory of the Military Technical Institute.

The activity concentrations of radionuclides were determined by high-resolution gamma spectrometry using a high-purity germanium (HPGe) detector with a relative efficiency of 56% and an energy resolution of 1.72 keV for the 1.33 MeV gamma line of ^{60}Co. Gamma-ray spectra were processed using the GammaVision 32 software package. The combined relative uncertainty associated with sample preparation and activity measurements did not exceed 10%.

The activity concentration of ^{134}Cs was determined using the gamma-ray emission lines at 604, 796, and 802 KeV, while ^{137}Cs was quantified using the 661.6 keV gamma line. The activity concentration of ^{40}K was determined from its characteristic gamma-ray emission at 1460 KeV.

3.1 Equipment and calibration

Calibration of the detector for measuring plant samples was done with a reference radioactive material, a matrix of silicone resin in Marinelli geometry. As a calibration source, a 1000ml mix standard of the MBSS 2 type in Marinelli geometry of the Inspectorate for ionizing radiation Czech metrological institute was used - certificate number: Cert.No: 9031-OL-159/08 total activities 41.48 kBq on August 31, 2023 (^{241}Am, ^{109}Cd, ^{139}Ce, ^{57}Co, ^{60}Co, ^{203}Hg, ^{88}Y, ^{113}Sn, ^{85}Sr ^{137}Cs). The time of measuring the samples is 60000 s.

The resolution depends on the energy deposited in the detector, that is, at higher energies, the resolution is better.

The first semiconductor material used for the detection of gamma radiation was silicon (Si), and later germanium (Ge) was used, which had an exceptional energy resolution, which is why it was widely used in spectrometry of photon radiation.In order to improve the performance of germanium detectors, a new material known as high-purity germanium with an active volume about 1 cm wide was obtained. Due to germanium detectors, gamma spectrometry is a widely used measurement technique for determining the concentration of radionuclide activity in various samples.

3.2 Sample Preparation

The selected food samples include cheese, eggs, butter, fruit, vegetables, grains, etc. A total of 20 samples of imported food brands were randomly sourced from local markets. The sample preparation procedure was strictly based on the recommendations of the International Atomic Energy Agency (IAEA). The measurement geometry was chosen based on the sample type; a plastic Marinelli beaker with a volume of 1000 ml was used for non-mineralized plant and meat samples. **[3]**

Table 3. Samples investigated in this stady:

Type	Code	Activity Concentration (Bq/kg)			
		^{137}Cs	^{238}U	^{232}Th	^{40}K
cheese	F 1	0,19± 0.03	0.041±0.006	0.026±0.004	56.4± 2.65
eggs	F 2	0.09± 0.04	0,46±0.02	0.14±0.02	9.59.± 0.68
butter	F 3	0.10±	0.04±0.01	0.03±0.01	35.23±

		0.02			3.45
hot dog	F 4	0.54± 0.06	0,34±0.02	0.16±0.06	12.66± 1.68
fruit	F 5	0.09± 0.02	0.046±0.0 06	0.031±0.0 05	98,2± 3.68
vegetabl es	F 6	0.40± 0.05	0.25,±0.03	0.18±0.02	143.23 ±4.32
corn grits	F 7	0,31± 0.02	0,62±0.04	0.46±0.02	141.2± 3.2
milk powder	F 8	0.1±0. 02	0.04±0.02	0.16±0.01	56.2± 3.2
cocoa	F 9	1,02± 0.07	2.48±0.11	1.58±0.08	832,4± 6.28
cottae cheese	F 10	0,21± 0.05	0.035±0.0 04	0.012±0.0 02	49.4± 1.73
rice	F 11	0.34± 0.04	1,34±0.01	0.40±0.02	65.8± 3.10
bacon	F12	0.21± 0.03	0.083±0.0 2	0.052±0.0 07	128.66 ±4.2
beans	F13	0.53± 0.04	6,35±0.05	0.74±0.06	138.5± 3.49
sardines	F14	0,78± 0.02	3,75±0.02	0.16±0.02	158.43 ±4.12
flour	F15	0,42± 0.04	0,81±0.05	0.61±0.04	122.66 ±5.2
salt	F16	<0.07	0.06±0.01	0.03±0.02	11.66± 2.3
sugar	F17	<0.05	0,02±0.01	0,02±0.01	6.13± 1.4
tea	F18	0.14± 0.02	0.13±0.02	0.22±0.01	35.66± 3.2
green salad	F19	0.37± 0.05	0.08±0.02	0.17±0.02	185.63 ±6.23
berries	F20	0.28± 0.05	0.08±0.02	0.06±0.02	116.24 ±8.45

3.3 Determination of radionuclide activity

Activity Concentration

The activity concentrations of the radionuclides in the samples are calculated from the net area of the corresponding photopeak using the following equation:

$$A = \frac{N_{net}}{\varepsilon \times \gamma \times m \times t_s}$$

– *Nnet* – The net area under the given photopeak of the corresponding gamma energy in the spectrum, calculated as:

$$N_{net} = N_s - N_b \times \left(\frac{t_s}{t_b}\right)$$

Ns – The total area of the photopeak in the sample spectrum.
Nb – The corresponding total area of the photopeak in the background radiation spectrum.
tb – The total time of the background radiation spectrum measurement (s).
ts – The total time of the sample spectrum measurement (s).
ε – The detection efficiency at the specific gamma-ray energy.
γ – The absolute transition probability of the gamma-ray emmision
m – The mass of the sample (kg).

3.4 Assessment of radiological risk from food ingestion

The assessment of radiological risk associated with the consumption of contaminated food is based on internal radiation exposure, as the ingestion of radionuclides results in their distribution within human tissues and the consequent internal irradiation of the body. The annual effective dose for an individual consumer was calculated using the standard equation recommended by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR).

$$E(mSvy^{-1}) = \sum(A_i \times U \times D_{f,i}) \times 10^{-3}$$

Where:

- ***E*** *– annual effective dose due to food ingestion (mSv·y⁻¹)*
- ***Ai*** *– specific activity concentration of the radionuclide (^{226}Ra, ^{232}Th, or ^{40}K) measured in food samples (Bq/kg);*
- ***$D_{f,i}$*** *– standard ingestion dose conversion factor for radionuclide (Sv/Bq);*
- ***10^{-3}*** *– conversion factor used to express the annual effective dose in millisieverts (mSv) rather than sieverts (Sv).*

The standard ingestion dose conversion factors (D_f) for adults, as recommended by the International Atomic Energy Agency (IAEA), represent the committed effective dose per unit intake of a radionuclide and are as follows:

- ^{40}K: 6.2×10^{-9} Sv·Bq⁻¹
- ^{226}Ra: 2.8×10^{-7} Sv·Bq⁻¹
- ^{232}Th: 2.3×10^{-7} Sv·Bq⁻¹

Table 4. Annual Effective Dose from Food Ingestion

Sample Type	Annual consumption U(kg/y)	Dose from ^{238}U (mSv·y⁻¹)	Dose from ^{232}Th (mSv·y⁻¹)	Dose from ^{40}K (mSv·y⁻¹)	Total anual Dose E (mSv·y⁻¹)
Cheese	15	1.72×10^{-7}	8.97×10^{-8}	5.25×10^{-6}	5.51×10^{-6}
Eggs	15	1.93×10^{-6}	4.83×10^{-7}	8.92×10^{-7}	3.31×10^{-6}
Butter	10	1.12×10^{-7}	6.90×10^{-8}	2.18×10^{-6}	2.36×10^{-6}
Hot dogs	20	1.90×10^{-6}	7.36×10^{-7}	1.57×10^{-6}	4.21×10^{-6}
Fruit	40	5.15×10^{-7}	2.85×10^{-7}	2.44×10^{-5}	2.52×10^{-5}
Vegetables	50	3.50×10^{-6}	2.07×10^{-6}	4.44×10^{-5}	5.00×10^{-5}
Corn grits	15	2.60×10^{-6}	1.59×10^{-6}	1.31×10^{-5}	1.73×10^{-5}
Milk powder	5	5.60×10^{-8}	1.84×10^{-7}	1.74×10^{-5}	1.98×10^{-6}
Cocoa	5	3.48×10^{-6}	1.82×10^{-6}	2.58×10^{-5}	3.11×10^{-5}
Cottage cheese	20	1.96×10^{-7}	5.52×10^{-8}	6.13×10^{-6}	6.38×10^{-6}
Rice	25	9.38×10^{-67}	2.30×10^{-6}	1.02×10^{-5}	2.19×10^{-5}
Bacon	15	1.39×10^{-7}	1.79×10^{-7}	1.20×10^{-5}	1.23×10^{-5}
Beans	15	2.67×10^{-5}	2.55×10^{-6}	1.29×10^{-5}	4.22×10^{-5}
Sardines	10	1.05×10^{-5}	3.68×10^{-7}	9.82×10^{-6}	2.07×10^{-5}
Flour	50	1.13×10^{-5}	7.02×10^{-6}	3.79×10^{-5}	5.62×10^{-5}
Salt	2	3.36×10^{-8}	1.38×10^{-8}	1.45×10^{-7}	1.92×10^{-7}
Sugar	10	5.60×10^{-8}	4.60×10^{-8}	3.80×10^{-7}	4.82×10^{-7}
Tea	2	7.28×10^{-8}	1.01×10^{-7}	4.42×10^{-7}	6.73×10^{-7}
Green salad	10	5.4×10^{-7}	2.76×10^{-6}	7.44×10^{-8}	6.16×10^{-7}
Sample 20	3	2.25×10^{-7}	1.15×10^{-6}	3.1×10^{-8}	3.09×10^{-6}

$$Ei = A \times U \times D_f$$

where:
Ei – annual effective dose resulting from the ingestion of radionuclide i (mSv·y⁻¹);
A – activity concentration of the radionuclide in food (Bq·kg⁻¹);
U – annual food consumption per person (kg/y);
Df – ingestion dose conversion factor for the radionuclide of interest (Sv/Bq).

4. RESULTS AND DISCUSSION

4.1 Determination of activity concentration

The activity concentrations of the analyzed radionuclides in the investigated samples of daily meals prepared at the Military Medical Academy are presented as mean measured values. Among the analyzed radionuclides, the naturally occurring radionuclide ^{40}K exhibited the highest activity concentrations in all food samples, ranging from 6.13 to 832.40 Bq·kg^{-1}, with a mean value of 120.19 Bq·kg^{-1}. The highest ^{40}K activity concentration was measured in the cocoa sample, which can be attributed to the naturally high potassium content of this food product.

The activity concentrations of ^{137}Cs were low in all analyzed samples and fell within the range typically reported for food products available on the commercial market. All measured values were below the recommended reference levels. The presence of this radionuclide reflects its long-term persistence in the environment following past nuclear accidents; however, only very low activity concentrations are currently detected in most food products. [4]

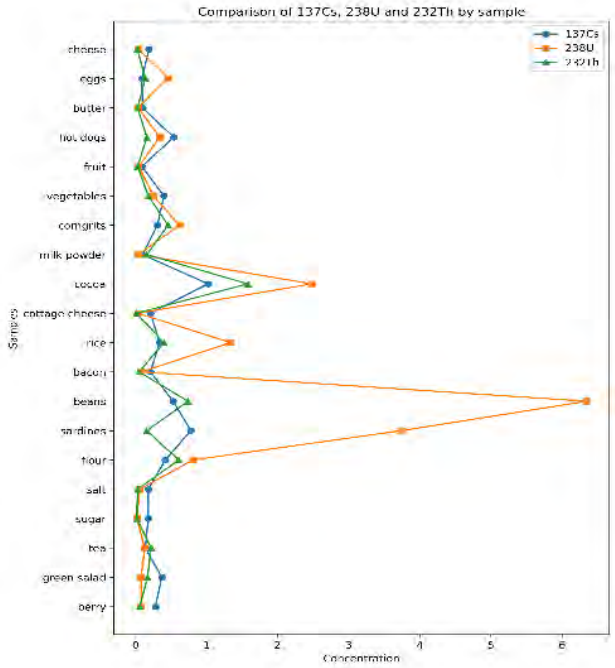


Figure 1. Comparison of Radionuclide Activity Concentrations among Food Groups

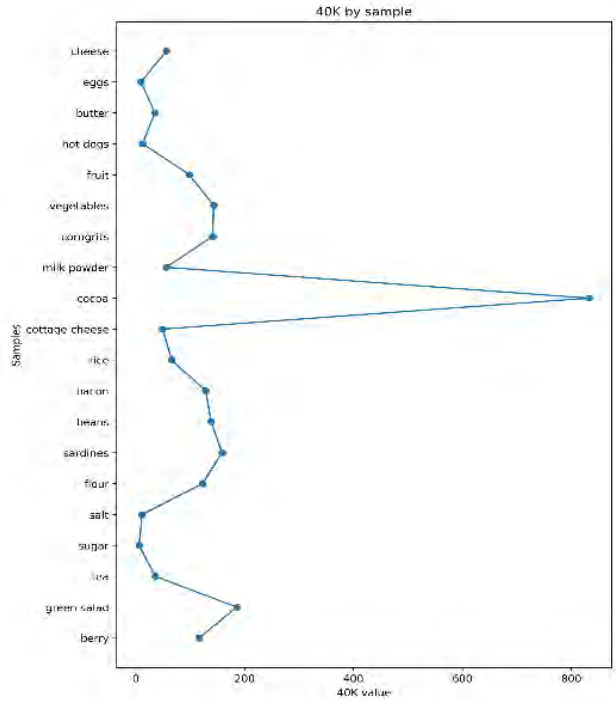


Figure 2. Comparison of the Specific Activity of ^{40}K among Food Groups

The concentrations of the natural radionuclides ^{238}U and ^{232}Th were considerably lower compared with ^{40}K. The highest uranium activity concentration was determined in the bean sample (6.35 Bq·kg^{-1}), while the highest thorium concentration was measured in the cocoa sample (1.58 Bq·kg^{-1}). Both radionuclides were detected at significantly lower levels than ^{40}K.

These results indicate that the influence of geological characteristics of the soil and the availability of specific minerals, represent dominant factors affecting the presence and transfer of radionuclides within the food chain. [5]

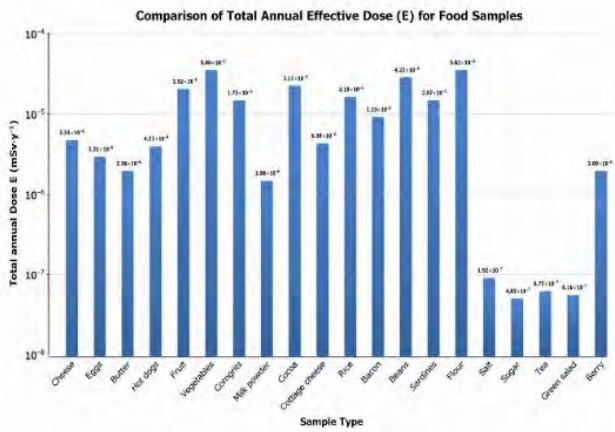


Figure 3. Comparison of the Total Annual Effective Dose for food samples

All obtained results were within the range of recommended values. The findings indicate that the food products analyzed for the assessment of the radiological burden of daily meals prepared at the Military Medical Academy do not represent a significant source of exposure of Serbian Armed Forces personnel to ionizing

radiation. The largest contribution to the total activity originates from the naturally occurring radionuclide ^{40}K, which is an essential component of the normal metabolism of living organisms and does not represent an indicator of radioactive contamination.

5. CONCLUSION

Gamma-spectrometric analysis of selected food products used in meal preparation showed that the activity concentrations of naturally occurring radionuclides (^{226}Ra,^{232}Th,and^{40}K) as well as anthropogenic radionuclide ^{137}Cs were within the ranges reported for similar food products in the available literature.

The estimated annual effective ingestion doses calculated from the measured activity concentrations indicate that consumption of the analyzed food products does not represent a significant source of radiological exposure for members of the Serbian Armed Forces. As expected, the largest contribution to the total annual effective dose originated from naturally occurring radionuclide ^{40}K reflecting its ubiquitous presence in food and its role in normal biological processes.

The obtained results confirm that the analyzed food groups are radiologically acceptable for consumption within the investigated dietary system and provide a reliable baseline for future monitoring of radionuclides in food products used in institutional nutrition.

References

[1] A.Feyisayo, I.Umar, M.Abdullahi, S.Dauda, Investigation of radioactivity levels in some selected imported canned foods consumed in Abuja Nigeria , J*ournal of radiation and nuclear application Vol 10.* No.1,pp.13-20 (2025)

[2] ORTEC. Food monitoring note: Radioactivity and its measurements in foodstufs,pp. l6-19

[3] UNSCEAR , *Sources and effect of ionizing radiation*, Unated Nations Scientific Committee on the Effects of Atomic Radiation, New York ,USA, pp.1-21 (2000)

[4] Direktorat za radijacionu i nuklearnu sigurnost i bezbednost Srbije; Godišnji izvestaj o stepenu izloženosti stanovništva jonizujućim zračenju u Republici Srbiji, (2024).

[5] International Atomic Energy Agency (IAEA), Generic models for use in assessing the impact of discharges of radioactive substances to the envoronment, Safety Reports Series No.19, IAEA ,Viena (2001)

[6] S. Forkapić. Istraživanje prisustva primopredajnih i antropogenih radionuklida u lancu ishrane i procena efektivne doze za stanovništvo, Doktorska disertacija, Prirodn-omatematički fakultet, Univeruitet u Novom Sadu, (2010)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF THERMOMECHANICAL TREATMENT ON THE GRAIN STRUCTURE OF Al-Mg-Mn ALLOYS FOR SHEAR SPINNING APPLICATION

LJUBICA RADOVIĆ
Military Technical Institute, Belgrade, ljubica.radovic@mod.gov.rs

MILJANA POPOVIĆ
University of Belgrade, Faculty of Technology and Metallurgy, miljana@tmf.bg.ac.rs

BRANKA JORDOVIĆ
University of Kragujevac, Faculty of Technical Science Čačak, branka.jordovic@ftn.kg.ac.rs

Abstract: *The commercial AlMg4.5Mn and AlMg6Mn alloys, containing approx. 0.50%Mn both, were subjected to thermomechanical treatments (TMTs) in order to produce different final microstructures with small or large grains. TMT I included: cold rolling with 50% reduction + annealing at 320°C/3h; TMT II included: two-stage annealing (565°C/6h + 450°C/4h), cold rolling with 20% reduction and final annealing. Final annealing for AlMg4.5Mn alloy was 280°C/3h, while for AlMg6Mn alloy was 350°C/3h. Microstructure was evaluated using optical microscopy and quantitative metallography. Microstructure after TMT I route was homogeneous for both alloys, with unimodal distribution around average grain size of 13-14 μm. TMT II route enabled the grain coarsening, but produced less homogeneous microstructure with a grain size in the range of 15-55μm, for both alloys. Final grain structure is attributed to interrelation between cold rolling reduction, chemical composition and annealing conditions.*

Keywords: *Thermomechanical treatment (TMT), Al-Mg-Mn alloys, Recrystallisation, Grain size, Grain size distribution*

1. INTRODUCTION

Aluminum-magnesium alloys are of interest in research for theirs many desirable properties, such as low density, high weight-to-strength ratio, high thermal and electrical conductivities, corrosion resistance, weldability, and relatively high ductility [1].

Aluminum-magnesium alloys sheets are widely used in the aerospace, automotive and ship-building industries and they are subject of continuous research. Military applications such as aircrafts, missiles, vehicle components, ships, boats, underwater weaponry, and light armor also require components that are strong and yet lightweight [2-5]. High strength, heat treatable, aluminum alloys (2xxx, 6xxx and 7xxx series) exhibit problems like consistency of tolerances introduced by quenching. Using non-heat treatable (solid solution hardened) Al-Mg alloys (5xxx series), this problem is eliminated, and additionally, these alloys have better corrosion resistance and good formability.

The strength of Al-Mg alloys is achieved mostly by solid solution strengthening and strain-hardening, and less by precipitation and grain boundary strengthening [6]. The main alloying element - magnesium, provides the most enhancement of strength among all alloying additions in aluminum solid solution, due to its high solubility in solid solution [1]. Hardness and strength of Al-Mg alloys linearly increase with addition of Mg [1]. Even an increase in Mg content above 5 wt% kept in good balance strength and formability [7]. Magnesium also, like the other solute elements, has a strong influence on recrystallization and grain growth, reducing both the rate of growth and the final grain size [8-9]. The addition of Mn in Al-Mg alloys, which have a limited solubility in solid solution, has a minor influence on strength of these alloys, but it has a significant influence on annealing behavior [10]. Mn containing particles usually precipitates during preheating, and inhibits the grain growth [11-12].

The chemical composition and processing parameters affect the development of the microstructure in these alloys. Grain size is known to be important for many properties of Al-Mg alloys, determining their mechanical properties, serrated flow and formability. Thus, controlling and evaluating grain size is important, especially for their formability [13-16]. The effect of applied TMTs on the grain size and its distribution in the past has been rarely discussed, particularly in Al-Mg-Mn alloys with high Mg content and Mn addition. A particular challenge is to increase the grain size due to the presence of Mn containing dispersoids on the grain boundary [17-19]. Additionally, there is lack of data on correlation between grain size and formability by spinning.

The aim of this work was to examine the influence of different thermomechanical treatments (TMTs) on the grain structure development in two Al-Mg-Mn alloys intended for further shear forming processing. This study

Corresponding author: LJ. Radović
E-mail: ljubica.radovic@mod.gov.rs
DOI: 10.5937/OTEH26128R

was performed in order to obtain fine grained (10-15 µm) and coarse grained structure (40-50 µm), and to evaluate grain size and its distribution that can affect further processing by shear spinning. The selected range of grain size was based on the preliminary research [20].

2. EXPERIMENTAL PROCEDURE

The chemical composition of the as received alloys is given in Table 1.

Table 1. Chemical composition of Al-Mg-Mn alloys

Alloy	Mg	Mn	Si	Cu	Fe	Zn	Ni	Al
AlMg4.5Mn	4.5	0.47	0.16	0.02	0.42	0.02	0.01	Bal.
AlMg6Mn	6.02	0.54	0.12	0.015	0.36	0.07	0.01	Bal.

Two TMTs were applied in an attempt to obtain microstructure with small (10 µm -15 µm) and coarse (40 µm -50 µm) grain, enable for a further sheet forming.

The as received hot rolled 15 mm and 8 mm thick plates of AlMg4.5Mn and AlMg6Mn alloys, respectively, were cold rolled up to 6 mm before TMT I, and up to 3,75 mm before TMT II.

TMT I route included: cold rolling (CR) to 50% reduction and annealing at 320°C/3h followed by air cooling (AC) for both alloys.

TMT II route included: two stage annealing (565°C/6h FC+450°C/4h FC to RT) (FC-furnace cooling, RT-room temperature). Following this process the samples of both alloys were cold rolled to 20 % reduction. Subsequent annealing was applied at 280°C/3h (AC) for AlMg4.5Mn and 350°C/3h (AC) for AlMg6Mn alloy.

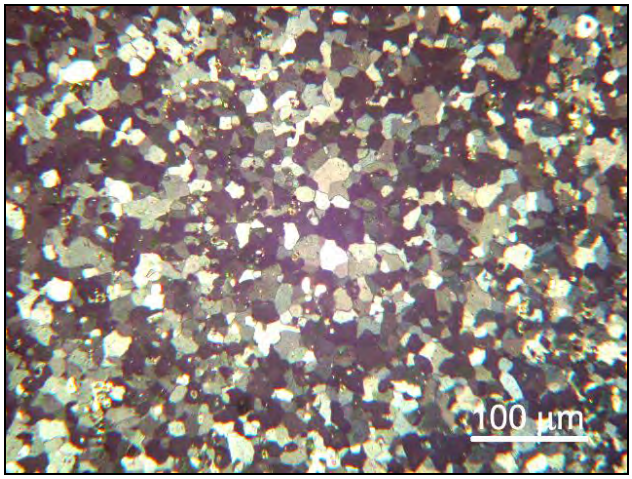
Thermomechanical treatments were determined on the base of published data and previous research [19, 21]. The final thickness of all specimens were 3 mm.

To reveal the grain structure, after electropolishing in perchloric acid, the samples were electrolytically etched in Barker′s solution. The microstructures were observed with a Leitz optical microscope (in polarized light) with a digital camera. Quantitative metallographic analysis was carried out using LAS-Leica Application Suite software V4.3 for image analysis. The grain size was determined by the linear intercept method.

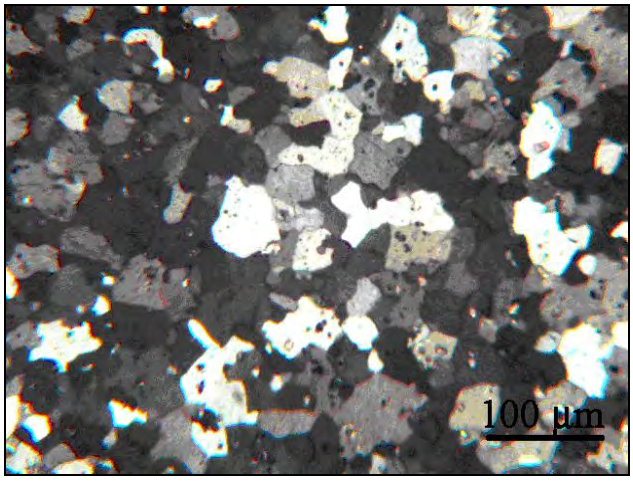
3. RESULTS AND DISCUSSION

Microstructures of Al-Mg-Mn alloys after applied two different TMTs are shown in the Figures 1 and 2. After TMT I microstructures of both alloys are characterized by small, equiaxed, recrystallized grains (Figures 1a and 2a). Observation by optical microscopy also revealed homogenous microstructure of small grain AlMg4.5Mn alloy (Figure 1a) compared to AlMg6Mn alloy (Figure 2a). After TMT II, which included two-stage annealing before cold rolling of 20 % reduction and subsequent annealing, the increase in grain size is achieved in both alloys (Figures 1b and 2b). However, microstructures are not homogenous, concerning the grain size.

Obtained grain size distributions for both examined alloys after applied TMTs are shown in Figures 3 and 4. The grain size distribution of AlMg4.5Mn and AlMg6Mn alloys with small grain are very similar (Figures 3a and 3b). The grain size distribution of AlMg6Mn alloy only exhibits a little longer tail in the coarse grain region compared to AlMg4.5Mn with the same grain size. The both distributions are unimodal, slightly asymmetric on the right. Also, the variation of the grain size is small. The dominant grain size is close to 15 µm for both alloys.



a)



b)

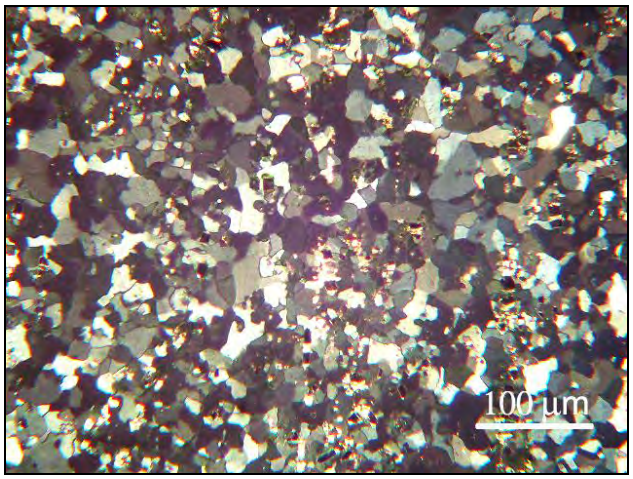
Figure 1. Microstructure of AlMg4.5Mn alloy: a) after TMT I: 50 % CR+320°C/3h AC; b) after TMT II: (565°C/6h FC+450°C/4h FC) + 20 % CR + 280°C/3h.

Cumulative curves show that approximately 90 % and 80 % of grains in small grain sized AlMg4.5Mn and AlMg6Mn alloy, respectively, are up to 20 µm. These results confirmed the microstructure homogeneity observed by optical microscopy.

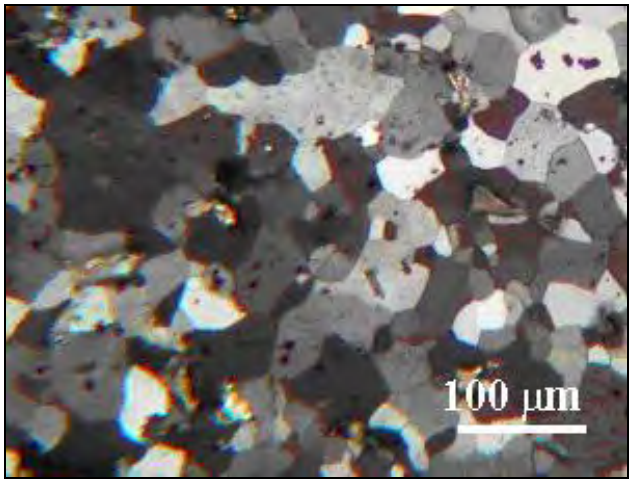
The grain size distribution of examined alloys after TMT II is shown in Figures 4a and 4b. It was shown that the range of the grain size in both alloys is large, larger than in small-grained specimens. Also, there was no single dominant grain size value. Cumulative curves show that approximately 75 % of grains are between 20 and 55 µm in AlMg4.5Mn alloy, i.e. 75 % between 15 and 55 µm in AlMg6Mn alloy.

Since the variation of the grain sizes in the microstructures obtained after TMT I is small, the average

grain size can represent these microstructures. Results of grain size quantification of Al-Mg-Mn alloys after different TMTs are shown in Table 2.



a)



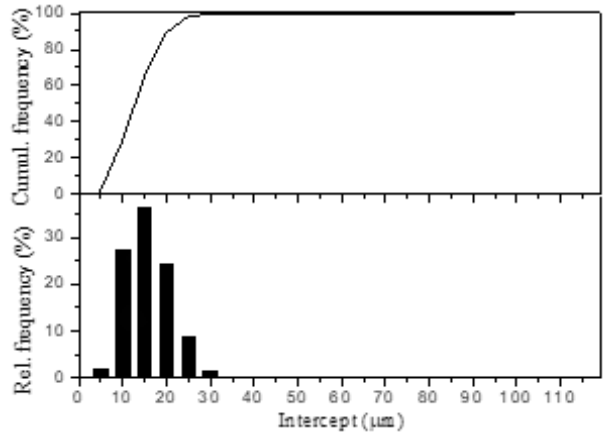
b)

Figure 2. Microstructure of AlMg6Mn alloy: a) after TMT I: 50 % CR+320°C/3h AC; b) after TMT II: (565°C/6h FC+450°C/4h FC) + 20 % CR +350°C/3h

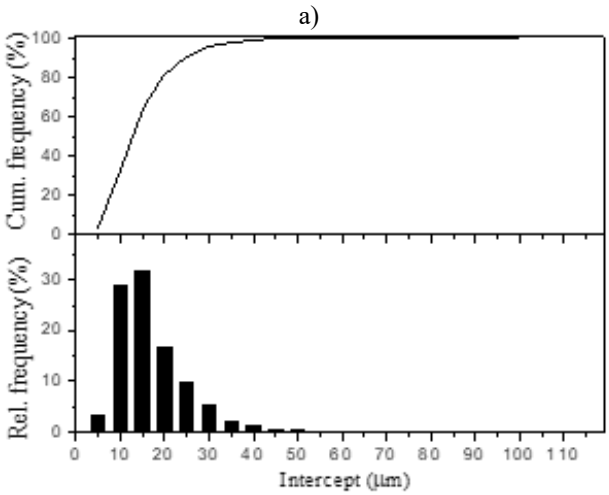
Table 2. Grain size in Al-Mg-Mn alloys after different TMTs.

Alloys	d_{av} (µm)	Range of the grain sizes, (µm)
	Small grained (TMT I)	Coarse grained (TMT II)
AlMg4,5Mn	13,3 ± 5.8	20-55
AlMg6Mn	13.7 ± 7.8	15-55

The results show that very close average grain size is obtained in both alloys after the same TMT. The average grain size of AlMg4.5Mn alloy is 13.3 ± 5.8 µm and AlMg6Mn alloy is 13.7 ± 7.8 µm. However, in the case of coarse grain alloys, the average grain size was not determined, because of wide range of the grain sizes. In this case, the average grain size does not represent the distribution of the grain size well. Therefore, the grain size in these microstructures is ranged from min to max value. The results shown that about 75 % of all grains in AlMg4.5Mn are in the range of 20-55 µm, i.e. in the range of 15-55 µm in AlMg6Mn alloy. It was shown earlier that microstructure development after cold rolling and annealing of Al-Mg alloys is closely related to both deformation and annealing temperature [11, 22-23]. The more homogenous microstructure in small grained specimens was obtained (Figures 1 and 2). The recrystallized grain size depends on the nucleation rate and growth rate. The effect of cold rolling reduction on recrystallized grain size is a direct result of an increase in the density of nucleation sites in the deformed microstructure. So, the final grain size in fully recrystallized small grain alloys is a result of an increase in the density of nucleation sites due to high reduction (50 %) before final annealing [14-16]. It also resulted in uniform grain growth, what was confirmed by grain size distribution results (Figure 3).



a)



b)

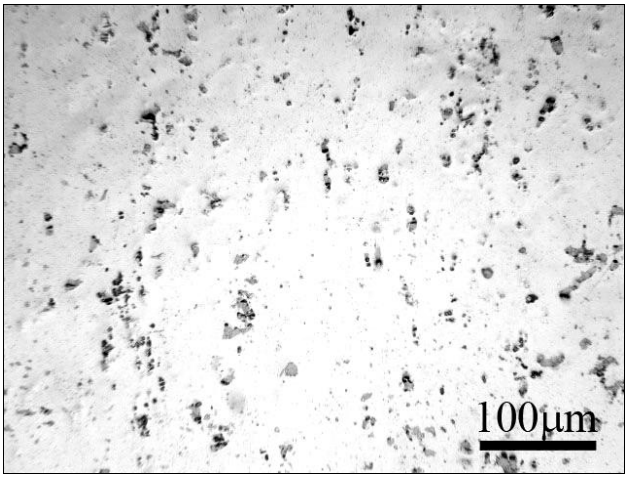
Figure 3. Grain size distribution in alloys with small grain (TMT I): a) AlMg4.5Mn, b) AlMg6Mn.

The grain growth after recrystallization in Al-Mg alloys is not significant, particularly with Mn addition [21, 27].

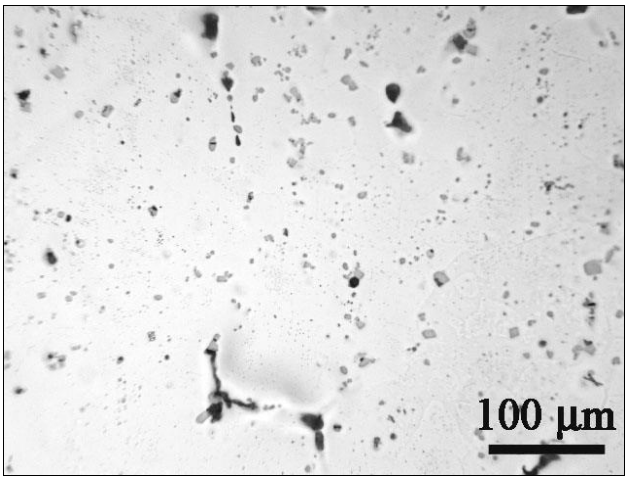
However, it was shown that in studied Al-Mg alloys with Mn addition, the grain growth during static recrystallization did not occur at critical strain (20 % and 25 % for AlMg4.5Mn and AlMg6Mn, respectively) and at 320°C [28]. Even more, the grain growth did not occur at increased temperature up to 500°C [29]. It was supposed that Mn bearing dispersoids (Al_6Mn) at grain boundaries

decreases the grain boundaries mobility during recrystallization and prevent the grain growth [1, 10, 18-19]. In order to alter the size and distribution of the second phase particles, two-stage annealing was applied. At first stage of annealing the fine, soluble second phase particles can be dissolved as much as possible, and then at second stage, during slow cooling rate and by keeping alloy at lower temperature, accelerate the growth of coarser particles [18]. It resulted in a decrease of the dispersoids density, interparticle distance increased, as published earlier for Al-Mg alloys [30]. It allowed the mobility grain boundary and grain growth.

Obviously, size and distribution of the second phase particles played a key role in grain size control and microstructure homogeneity. Increases of the particle size and interparticle space led to grain growth. The differences in size and particle distribution after TMT I and TMT II were observed in both alloys. Figure 4 shows particle distribution in AlMg6Mn alloy.



a)

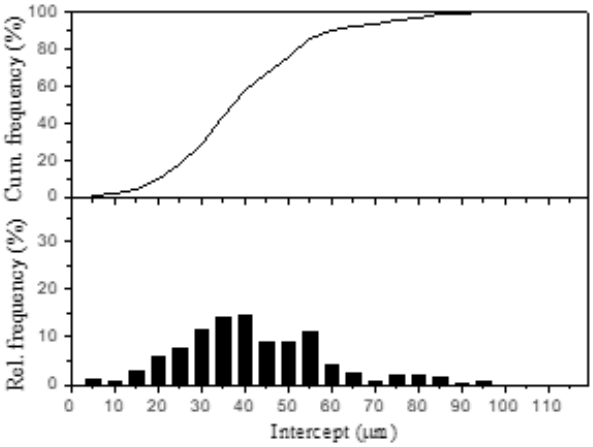


b)

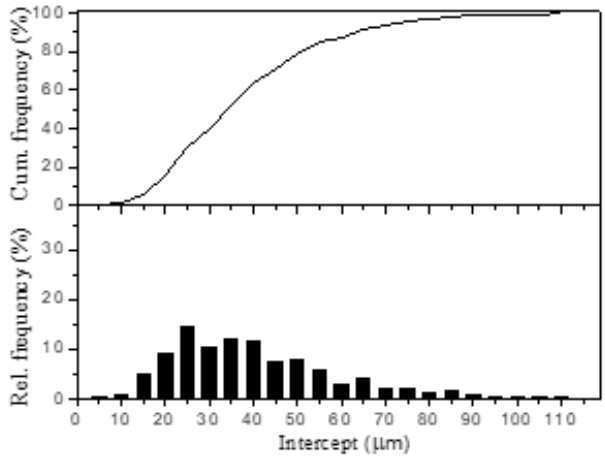
Figure 4. Particle distribution in AlMg6Mn alloy: a) TMT I): b) TMT II.

It was assumed that the inhomogeneous microstructure after recrystallization during TMT II is result of inhomogeneous distribution of the second phase particles and low reduction (e=20-40 %) before final annealing [18, 31]. In such microstructure the average grain size is not convenient for microstructure characterization, so the range of the grain sizes is given (Table 2).

The grain size distribution in both alloys after TMT II is shown in Figure 5.



a)



b)

Figure 5. Grain size distribution in alloys with coarse grained structure (TMT II): a) AlMg4.5Mn; b) AlMg6Mn.

The distribution of the grain size in coarse grain alloys shows the wide range of the grain size. It is supposed that such behavior can be result of inhomogeneous distribution of the second phase particles and small reduction (20 %) before final annealing.

4. CONCLUSION

After TMT which included 50 % CR+320°C/3h AC, homogenous microstructure in AlMg4.5Mn and AlMg6Mn alloys obtained. The distribution of the grain size is unimodal, and the average grain size was 13.3 µm and 13.7 µm in AlMg4.5Mn and AlMg6Mn alloy, respectively.

High temperature two-stage annealing (565°C/6h FC+450°C/4h FC) followed by 20 % reduction and final annealing at 280°C/3h for AlMg4.5Mn and 350°C/3h for AlMg6Mn, resulted in coarser grains in both alloys.

About 75 % of all grains are in the range of 20 µm to 55 µm in AlMg4.5Mn, i.e. 15 µm to 55 µm in AlMg6Mn alloy.

The less homogenous microstructure in coarse grain Al-Mg-Mn alloys is attributed to inhomogeneous distribution of the second phase particles after two-stage annealing and small reduction (20 %) before final annealing.

Acknowledgement

"This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contracts No.451-03-34/2026-03/200325 and No. 451-03-34/2026-03/ 200135).

References

[1] MONDOLFO, L. F., Aluminium Alloys: Structure and Properties, Boston, Butterworths, 1976

[2] TEMPUS, G., New aluminium alloys and fuselage structures in aircraft design, Materials Day, “Werkstoffe für transport und verkehr” ETH Zürich, Switzerland, 2001.

[3] BØRVIK, T., Forrestal, M.J., Warren, T.L., Perforation of 5083-H116 Aluminum Armor Plates with Ogive- Nose Rods and 7.62mm APM2 Bullets, Exp. Mech. 50 (2010) 969-978.

[4] ZHOU, T., et al. Investigation of the Al alloy armor materials: A review, Journal of Physics: Conference Series 2891 (2024) 162019.

[5] Application of High-End Aluminum Alloys in Military Industrial Materials, Longkou Mat Aluminum CO., LTD, mart 2025., access April 2026.

[6] HUSKINS, E.L., CAO, B., RAME, K.T., Strengthening mechanisms in an Al–Mg alloy, Mater. Sci. Eng. A 527 (2010) 1292–1298.

[7] ROMHANJI, E., POPOVIĆ, M., GLIŠIĆ, D., MILENKOVIĆ V., Formability of a high-strength Al-Mg6.8 type alloy sheet, J. Mater. Sci. 33 (1998) 1037-1042.

[8] Sanders Jr. R.E., Baumann S.F., Stumpf H.C., Aluminum Alloys-Phisical and Mechanical Properties, vol.III, EMAS, West Midlands, UK (1986) pp. 1441-1484

[9] RYUM, N., EMBURY, J.D., A comment on the recrystallization behaviour of Al-Mg alloys, Scand. J. Metall., 11 (1982) 51-54.

[10] LEE, S. L., Wu, S. T., Identification of Dispersoids in Al-Mg Alloys Containing Mn, Metall. Trans. A 18A (1987) 1353-135.

11] KANNAN, K., VETRANO, J.S., HAMILTON, C.H., Effects of Alloy Modification and Thermomechanical Processing on Recrystallization of Al-Mg-Mn Alloys, Metall. Mater. Trans. A, 27 (1996) 2947-2957.

[12] COURT, S.A GATENBY, K.M., LLOYD, D.J., Factors affecting the strength and formability of alloys based on Al–3 wt.% Mg, Mater. Sci. Eng. A, 19–321, 2001, 443-447.

[13] WEN, W., MORRIS, J.G., The effect of cold rolling and annealing on the serrated yielding phenomenon of AA5182 aluminum alloy, Mater. Sci. Eng. A 373 (2004) 204–216.

[14] WAGENHOFER, M., ERICKSON-NATISHAN, M. A., ARMSTRONG, R. W., Influences of Strain Rate and Grain Size on Yield and Serrated Flow In Commercial Al-Mg Alloy 5086, Scripta Mater. 41 No. 11 (1999) 1177-11841.

[15] CHENG, T.Ch., SHEAN LEE, R., The influence of grain size and strain rate effects on formability of aluminium alloy sheet at high-speed forming, J. Mater. Process. Technol. 253 (2018) 134–159.

[16] LLOYD, D. J., Court, S. A., Influence of grain size on tensile properties of Al –Mg alloys, Mater. Sci. Technol. (2003) 19, 1349-1354.

[17] LEE, S-L., Wu, S. T., Indentification of Dispersoids in Al-Mg Alloys Containing Mn, Metall. Trans. A Vol.18A (1987) 1353-1357.

[18] OSMAN. M., ENGLER, O., KARHAUSEN, K., LOCHTE, L., MCLARE, A.J., Effect of homogenisation conditions on recrystallisation in Al-Mg-Mn alloy AA 5454, Mater. Sci. Technol. 23 (6) (2007) 688-698.

[19] ZAIDI, M.A., SHEPPARD, T., Effect of high temperature soak and cooling rate on recristallization behaviour of two Al-Mg alloys (AA5252 and AA5454), Mater. Technol. 11 (1984) 313-319.

[20] RADOVIĆ, LJ., Deformation behaviour and fracture characteristics of Al-Mg alloys: composition-technology-microstructure correlation, PhD Thesis, Faculty of Technical Science, Čačak, 2013.

[21] RADOVIĆ, LJ., NIKAČEVIĆ, M., POPOVIĆ, M., ROMHANJI E., The influence of thermomechanical treatment and chemical composition on recrystallization of Al-Mg alloy, MJoM, No. 4, 13 (2007) pp. 259-267.

[22] RADETIC, T., POPOVIĆ, M., ROMHANJI, E., Influence of cold-rolling deformation and annealing temperature on the grain growth of Al-4%Mg-Mn alloy, Materiali in tehnologije/Mater. Technol. 54 (2020) 1, 85–90.

[23] RADOVIC, Lj., NIKAČEVIĆ, M., Microstructure and properties of cold rolled and annealed Al-Mg alloys, Scientific Technical Review, LVIII. No 2, (2008) 14-20.

[24] HUMPHREYS, F. J., A Unified Theory of Recovery, Recrystallization and Grain Growth, Based on The Stability and Growth of Cellular Microstructures-I. The Basic Model, Acta Mater. 45(10) (1997) 4231-4240.

[25] BIROL, Y., Effect of solute Mg on grain size of aluminium alloys, Mater. Sci. Technol. 28(8) (2012) 924-927.

[26] RAGHUNATHAN, N., SHEPPARD, T., Microstructural development during annealing of hot r*olled Al-Mg alloys*, Mater. Sci. Technol. 5 (1989) 542-547.

[27] GO, J., POOLE, W. J., MILITZER, M., Wells M. A., Modelling recovery and recrystallisation during annealing of AA 5754 aluminium alloy, Mat. Sci. Technol., 19 (2003) 1361-1368.

[28] RADOVIĆ, LJ., NIKAČEVIĆ, M., Effect of cold rolling reduction on critical strains and final grain size of Al-Mg alloys containing Mn, DPSM, ed. by E. Romhanji, M.T. Jovanović, N. Radović, Belgrade (2005) 231-236.

[29] RADOVIĆ, Lj., Thermomechanical treatment of Al-Mg alloys for shear spinning, Technical report, Military Technical Institute, (2005).

[30] LEE, S.-L., Wu, S.T., Influence of Soaking Treatments on Hot Ductility of Al-4.85 Pct Mg Alloys Containing Mn, Metall. Trans. A, 17A (1986) 833-841.

[31] JENSEN, J. D., Deformation Microstructure and Recristallization Behaviour, The fourth Int. Conf. on Recrystalliyation and Related Phenomena, Ed. by T. Sakai and H.G. Suzuki, The Japan Institute of Metals (1999) 3-14.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

DIFFERENT SYNTHETIC ROUTES OF CERIUM-DIOXIDE NANOPARTICLES PREPARATION AND THEIR EFFICIENCY IN SULFUR MUSTARD REMOVAL

MAJA VITOROVIĆ-TODOROVIĆ
Military Technical Institute, Belgrade, majavitod@gmail.com

ALEKSANDAR ĆITIĆ
Military Technical Institute, Belgrade, woteusrb@gmail.com

Abstract: *Highly toxic compounds such as chemical warfare agents (CWAs) are subject to strict international control, and their use is limited to protective and research purposes under national and international legislation, most notably the Chemical Weapons Convention. However, the potential use of these agents by terrorist groups, as well as their accidental release, continues to pose a serious threat to civilian populations. Ongoing efforts are focused on developing new strategies and technologies for the removal, decontamination, and detoxification of CWAs. One promising approach involves the use of metal oxide nanoparticles for the catalytic degradation of CWAs, such as magnesium oxide, titanium dioxide, and zirconium hydroxide being the most extensively studied materials. More recently, cerium dioxide (CeO_2) nanoparticles have demonstrated considerable potential for CWAs degradation, due to their ability to readily switch between +3 and +4 oxidation states. The degradation efficiency of CeO_2 nanoparticles strongly depends on their morphology, aggregation state, and surface area. In this study, we explored several synthetic routes for CeO_2 nanoparticles preparation, including homogeneous hydrolysis, the oxalate route, precipitation with ammonium-carbonate and the solution combustion synthesis. The morphology and structure of the obtained particles were characterized using scanning electron microscopy (SEM), Raman spectroscopy and N_2 adsorption experiments. Their degradation efficiency toward HD was evaluated by gas chromatography–mass spectrometry. In most cases, CeO_2 nanoparticles exhibited highly efficient removal of HD agent.*

Keywords: *cerium-dioxide, nanoparticles, catalytic degradation, chemical warfare agents*

1. INTRODUCTION

Chemical warfare agents (CWAs) belong to a group of exceptionally hazardous toxic chemicals whose production, possession and use are rigorously regulated by national legislation and international agreements, particularly the Chemical Weapons Convention. Although their legitimate handling is restricted to defensive, analytical and research activities, the risk associated with their intentional deployment in armed conflicts or terrorist attacks has not been eliminated. In addition, accidental releases during storage, transport or handling may also result in severe consequences for both human health and the environment. These concerns have stimulated continued research into advanced decontamination approaches, including more effective reactive formulations, improved detoxification technologies and the design of novel materials for individual and collective protection. In this sense, nanotechnologies, particularly the utilization of catalytically active metal-oxide nanoparticles as carriers of active chemistry represent a good starting point for the development of various decontamination formulations and novel protective materials. Stoichiometric nanocrystalline MgO was one of the first oxides proposed for the destruction of CWA [1], followed by other types of nanoparticles, including calcium-oxide (CaO) [2], titanium-dioxide (TiO_2) [3] and zirconium(IV)-hydroxide ($Zr(OH)_4$)[4]. While earth-alkaline oxides and $Zr(OH)_4$ decompose organophosphate agents *via* hydrolytic pathways, TiO_2 degrade organic substances by free-radical mechanism generated with water as an intermediate reagent in decomposition reactions. Here should be emphasized that not all CWAs are equally resistant toward degradation. For instance, G-class organophosphate agents namely soman (3,3-dimethylbutanyl methylphosphonofluoridate, GD) and sarin (2-propanyl methylphosphonofluoridate, GB) easily degrade by changing pH of the environment with the half-time of reaction being only 2 minutes in the pH above 8 and below 4. For detoxification of these agents, only hydrolytic cleavage of P-F bond is needed (Figure 1). On the contrary, some other types of agents are more resistant toward decontamination, and require both oxidative and hydrolytic degradation pathways. These are blister agent, sulfur mustard (1-chloro-2-[(2-chloroethyl)-sulfanyl]-ethane, HD) and VX (S-{2-[di(propan-2-yl)amino]ethyl} O-ethyl methylphosphonothioate). Degradation pathways of these two agents should be strictly controlled since there is a possibility of occurrence of byproducts which retain their toxic properties.

Corresponding author: M. Vitorović-Todorović
E-mail: majavitod@gmail.com
DOI: 10.5937/OTEH26129V

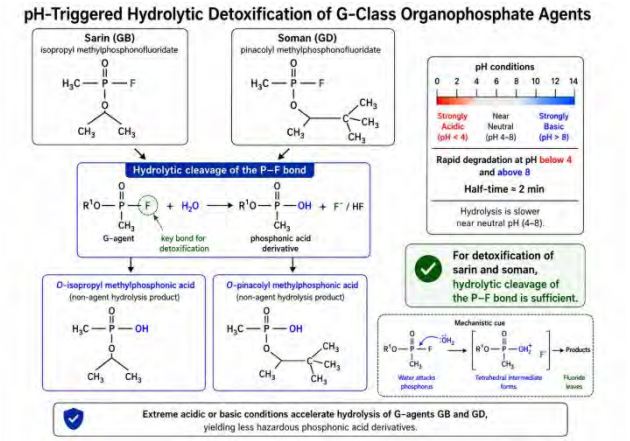


Figure 1. Hydrolytic detoxification of G-agents.

For instance, the partial oxidation pathway of sulfur mustard should be strictly controlled to yield sulfoxide product only, since further oxidation to sulfone gives product with blistering properties (Figure 2) [5].

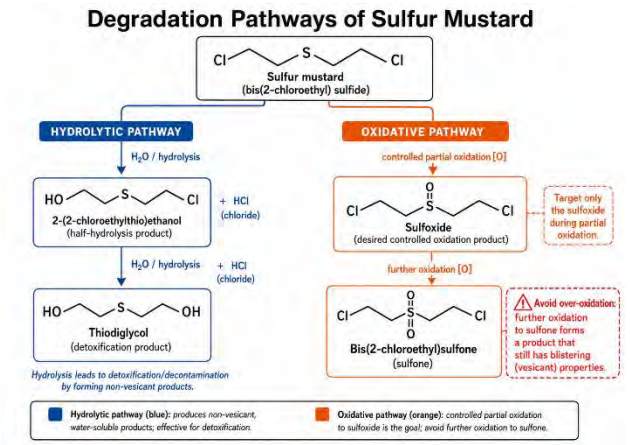


Figure 2. Degradation pathways of sulfur mustard.

The hydrolytic pathways of VX should be also strictly selective in a way that yield product of P-S bond cleavage only. If the P-O bond is cleaved, a very toxic EA2192 product, with potent anticholinesterase activity is obtained [6]. Moreover, further oxidation of diisopropylethylthio moiety is desirable (Figure 3).

Based on all of these stated above, there is a need for efficient decontamination agent which may provide both oxidative and hydrolytic pathways of CWA degradation. In this sense, nanoparticles of cerium-dioxide (CeO_2) seemed as a promising agent due to the several interesting properties. Cerium is the first element of lanthanide series. Its 4f orbitals are shielded by 5p and 4d orbitals, the electronic configuration of Ce+3 ion is $[Xe]4f^1$. By losing 3 electrons to achieve a +3 oxidation state, it gives up the two 6s and one 5d electron leaving only one 4f electron.

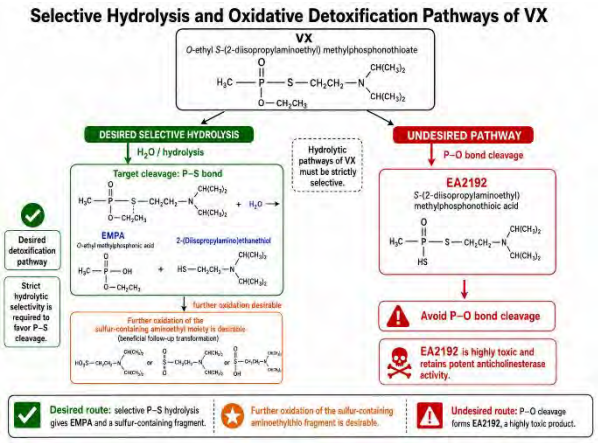


Figure 3. Degradation pathways of VX.

Due to an interesting electronic configuration, Ce exists easily in 3+ and 4+ oxidation states, meaning that $CeO_2 \rightarrow Ce_2O_3$ transition easily takes place, especially if the material is in its nano-form. Therefore, nanoceria has significant capacity to store and release oxygen and repeatedly pass through redox cycles rapidly. Due to these particular properties, CeO_2 nanoparticles are widely used as fuel oxidation catalyst [7], automotive exhaust treatment [8] and also these nanoparticles may exhibit enzyme mimicking activities including superoxide-dismutase, catalase, and phosphatase enzymes [9], [10].

The degradation efficacy of CeO_2 nanoparticles is strongly dependent upon their crystallite size, presence of oxygen vacancies and the presence of crystal lattice defects as well as their morphology. On the other hand, all these properties depend upon the synthetic routes used to obtain CeO_2 nanoparticles. In this study, we examined the influence of the synthetic routes for CeO_2 on their morphology and degradation efficacy using sulfur mustard. The aim of this work was to decipher whether degradation efficacy may be finely tuned by the synthetic routes used to obtain CeO_2 nanoparticles and how their morphology is related to their catalytic activity. For the synthetic routes we have chosen the most employed and the best described in the scientific literature, including: homogenous hydrolysis, precipitation with oxalic acid, precipitation with ammonium-bicarbonate and solution combustion synthesis, using urea as a fuel. Together with commercially available CeO_2 nanoparticles (25 nm in size, Sigma Aldrich) amorphous and the same that is annealed at 500 °C to induce crystallization. This yielded totally six different CeO_2 nanoparticles samples (Table 1).

The morphology and structure of the obtained particles were characterized using scanning electron microscopy (SEM), Raman spectroscopy and N_2 adsorption (BET isotherm analysis). Their degradation efficiency toward HD was evaluated by gas chromatography–mass spectrometry.

Table 1. CeO_2 nanoparticle samples investigated in this work.

Sample code	Synthetic route / origin	Key reagent or treatment	Brief description
CeO_2-1	Homogenous hydrolysis	Urea	CeO_2 nanoparticles obtained by homogeneous hydrolysis
CeO_2-2	Precipitation	Oxalic acid	CeO_2 nanoparticles obtained via oxalate precipitation route
CeO_2-3	Precipitation	Ammonium bicarbonate	CeO_2 nanoparticles obtained by carbonate/bicarbonate precipitation
CeO_2-4	Solution combustion synthesis	Urea as fuel	CeO_2 nanoparticles synthesized by solution combustion route
CeO_2-5	Annealed commercial CeO_2	Annealed at 500°C	Commercial CeO_2 nanoparticles, annealed at 500°C
CeO_2-6	Commercial CeO_2	Sigma-Aldrich, 25 nm	Commercial CeO_2 nanoparticles

2. MATERIALS AND METHODS

All chemicals used in this work were purchased from Merck or Sigma Aldrich and were used without further purification. The HD agent was synthesized at Military Technical Institute, Belgrade and its purity was checked by the GC/MS system, confirming at least 95% of purity.

Caution: *Chemical warfare agents are highly toxic and strictly regulated substances. All CWA-related experiments were conducted solely for the research purposes in an authorized facility by trained personnel, following approved safety, containment, decontamination, and waste-disposal procedures, in compliance with institutional, national, and Chemical Weapons Convention requirements. Unauthorized handling or use of these substances must not be attempted.*

2.1. CeO_2 nanoparticles preparation

Homogenous hydrolysis

The 3 g (6.91 mmol) of $Ce(NO_3)_3x6H_2O$ were placed in the 50 mL beaker and dissolved in 10 mL of hot deionised water with the addition of 1,5 mL of concentrated nitric acid. This solution was transferred into 1L round bottom 2- or 3-necked flask equipped with a magnetic stirrer and a reflux condenser. This solution was diluted with the 400 mL of deionised water, and 50 g of urea was added divided in ten smaller portions with stirring applied. After that, the mixture was heated at 90°C, to induce slow and steady decomposition of urea and consequent increase in pH of the solution, which causes white precipitate of $Ce(OH)_4$ to occur. After cooling the mixture to room temperature, the precipitate was separated by filtration and dried in the atmosphere at 105 °C. The $Ce(OH)_4$ obtained was further calcined at 500 °C for 2 hours to induce dehydration and crystallization of the CeO_2. This sample was designated as CeO_2_1.

Oxalate precipitation

The 8,6 g (19,8 mmol) $Ce(NO_3)_3x6H_2O$ were fully dissolved in 100 mL of deionized water, in 250 mL laboratory beaker that was placed on the magnetic stirrer. The 3 g of oxalic acid (23,88 mmol) was fully dissolved in the minimal amount of deionized water. This solution was transferred into dropping funnel and slowly, during 3h was added drop wise into solution of cerium-nitrate under vigorous stirring. The white precipitate of cerium(III)-oxalate is formed. The stirring of the mixture was continued for additional hour after the addition of oxalic acid. After that, the white precipitate was separated by filtration, and dried in the oven at 105 °C. The cerium-oxalate obtained was further calcined at 500 °C for 2 hours to induce combustion of the organic residue and crystallization of the CeO_2. This sample was designated as CeO_2_2.

Precipitation with ammonium-carbonate

The 8,6 g (19,8 mmol) $Ce(NO_3)_3x6H_2O$ were fully dissolved in 100 mL of deionized water, in 250 mL laboratory beaker that was placed on the magnetic stirrer. The 4,8 g of ammonium-carbonate (50,00 mmol) is fully dissolved in 100 mL of deionized water. This solution was transferred into dropping funnel and slowly, during 3h was added dropwise into the solution of cerium-nitrate under vigorous stirring. The white precipitate of cerium(III)-carbonate is formed. The stirring of the mixture was continued for additional hour after the addition of ammonium-carbonate. After that, the white precipitate was separated by filtration, and dried in the oven at 105 °C. The cerium-carbonate obtained was further calcined at 500 °C for 2 hours to induce combustion of the organic residue and crystallization of the CeO_2. This sample was designated as CeO_2_3.

The solution combustion synthesis

The stoichiometric fuel-to-oxidizer ratio was calculated according to the total oxidizing and reducing valences of the reactants, following the propellant chemistry approach. Considering urea as fuel with a reducing valence of +6 and cerium nitrate as an oxidizer with an effective oxidizing valence of −14 for CeO_2 formation, the stoichiometric molar ratio $Ce(NO_3)_3 \cdot 6H_2O$: urea was 1 : 7/3. Therefore, 4,3 g (9,9 mmol) of $Ce(NO_3)_3x6H_2O$ and 1,39 g (23.1 mmol) were dissolved in 30 mL of deionized water in a 250 mL beaker with vigorous stirring. After the full dissolution of the compounds, the mixture was heated slowly to evaporate a portion of water and to form a viscous gel. When the gel is being formed, the temperature of the heater was increased to 450 °C to induce combustion of the mixture. After combustion, the dry pale yellow CeO_2 is formed in the beaker. The cerium-dioxide obtained was further calcined at 500 °C for 2 hours to induce crystallization. This sample was designated as CeO_2_4.

Two other samples were: commercially available calcined at 500 °C was designated as CeO_2_5, and commercially available, amorphous without any thermal treatment was designated as CeO_2_6.

2.2. Morphological characterisation of the samples

The morphology and the size of the nanoparticles were investigated by scanning electron microscopy using JEOL

JSM-660 LV microscope. The surface of the samples was covered with layer of graphite to assure conductivity. The EDS analysis was done by energy dispersive spectrometer EDS OXFORDX-Max with Aztec software.

The Raman spectra of the samples CeO_2_2 and CeO_2_6 were recorded on Anton Paar Cora 5001 spectrometer. The BET surface areas for the samples CeO_2_3, CeO_2_5 and CeO_2_6 were determined by N_2 adsorption experiments using Anton Paar Nova 600 BET analyser.

2.3. HD degradation experiments

For the HD degradation experiments, the 30 mg of the each CeO_2 sample was placed in each of the eight eppendorf 2 mL microtubes. The solution of the HD of concentration 10 mg/mL was prepared in the *n*-hexane. The 50 μL of the HD solution was pipetted in shortest possible time interval in all test tubes. In the first tube, 950 μL of isopropanol was added immediately to terminate the degradation reaction (that is visibile almost as zero point time on the graph, while it is actually 20-30 seconds). After that, in the time intervals of 10, 20, 30, 40, 50, 60 and 70 min the isopropanol was added in test vials. The extraction of the remaining HD and possible degradation products was done by shaking the tubes for 15 minutes, and then the solid was separated by centrifugation at 12000 rpm. The 10 μL aliquot was drawn from the supernatant and placed in GC-MS vials filled with 980 μL of acetonitrile and 10 μL of the internal standard solution, naphthalene-d_8. The amount of HD was quantified and the resulted chromatogram inspected for the presence of degradation products. Calibration standards and the samples were analyzed on Shimadzu GC-MS QP2020 instrument, and the peak area in the chromatogram and the creation of the calibration curve (peak area vs. concentration) was done using Shimadzu LabSolution software. For the quantification of HD, the characteristic and the most intensive mass ion m/z 109 was used. The experimental conditions for the GC-MS instrument are given in the Table 2.

Table 2. GC-MS conditions

GC Conditions	
Parameter	**Value**
Starting column temperature	40 °C, 3 min
Temperature program	40-150°C, 10°C/min, 150-280 °C, 25°C/min, 280 °C, 10,8 min
Injector temperature	250 °C
Injection type	Grob type. Splitles t 0,75 min
Colum type	DB-5
Column dimensions	30 m x 0,25 mm ID x 0,25 μm
Carrier gas	Helium, 32 cm/sec
MS Conditions	
Parameter	**Value**
Temperature transfer line	280 °C
Temperature ion source	230 °C
Temperature detector	150 °C
Electron energy	70 eV
Mass range	35-500 *m/z*
Scanning rate	3,15 scan/sec
Library	NIST 05 Mass Spectral Data Base

3. RESULTS AND DISCUSSION

3.1. CeO_2 nanoparticles preparation

The all four synthetic procedures were easily reproduced, and the CeO_2 was obtained in relatively high yields. The differences between appearance of the crystallites for the different samples was obvious even with a naked eye. For instance, the samples obtained with precipitation by oxalic acid or ammonium-carbonate, CeO_2_2 and CeO_2_3 were very light and highly voluminous, while the sample obtained by solution combustion synthesis, CeO_2_4 was heavy, firm and appear more aggregated than the rest of the samples.

3.2. Morphology of the CeO_2 nanoparticles

The SEM images of all samples in two different magnifications are given in Figures 4 and 5.

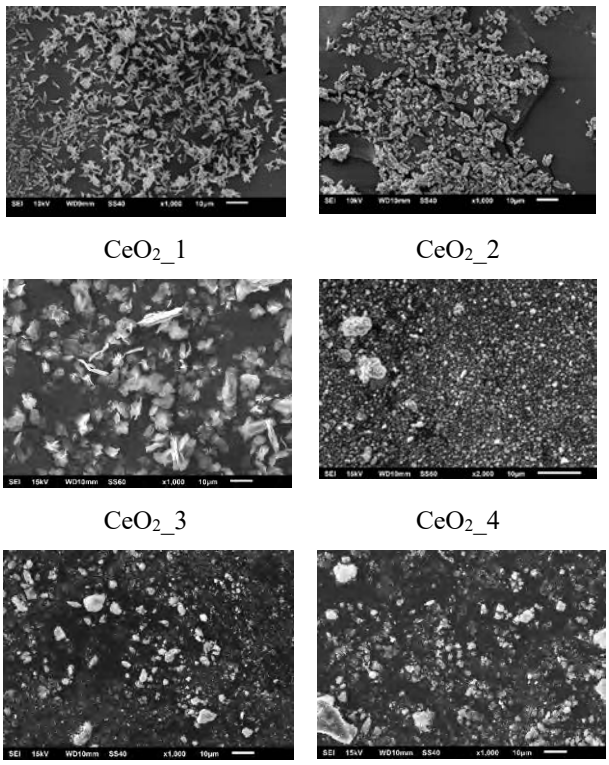


Figure 4. The SEM images of the CeO_2 samples under magnification of 1000x.

SEM analysis revealed pronounced morphology differences among the CeO_2 samples obtained by different synthetic routes. CeO_2_1, prepared by homogeneous hydrolysis, exhibited a highly anisotropic morphology composed of elongated rod-like and flower-like aggregates. CeO_2_2, obtained by oxalate precipitation, consisted mainly of irregularly shaped agglomerates with plate-like subunits. In contrast, CeO_2_3, synthesized by precipitation with ammonium carbonate, showed more open lamellar/flake-like morphology, with thin plate-like structures and exposed edges. CeO_2_4, prepared by solution combustion synthesis, displayed a finely dispersed granular morphology composed of small, quasi-spherical

particles forming porous aggregates. The annealed commercial sample, CeO_2_5, consisted of larger agglomerated particles, most likely due to thermally induced sintering, whereas the amorphous commercial CeO_2_6 showed strongly agglomerated, poorly defined particulate morphology.

The EDS analysis in all samples confirmed the presence of cerium and oxygen atoms only (results not shown).

The Raman spectra of CeO_2_2 sample (crystalline) and CeO_2_6 sample (amorphous) are given in Figure 6.

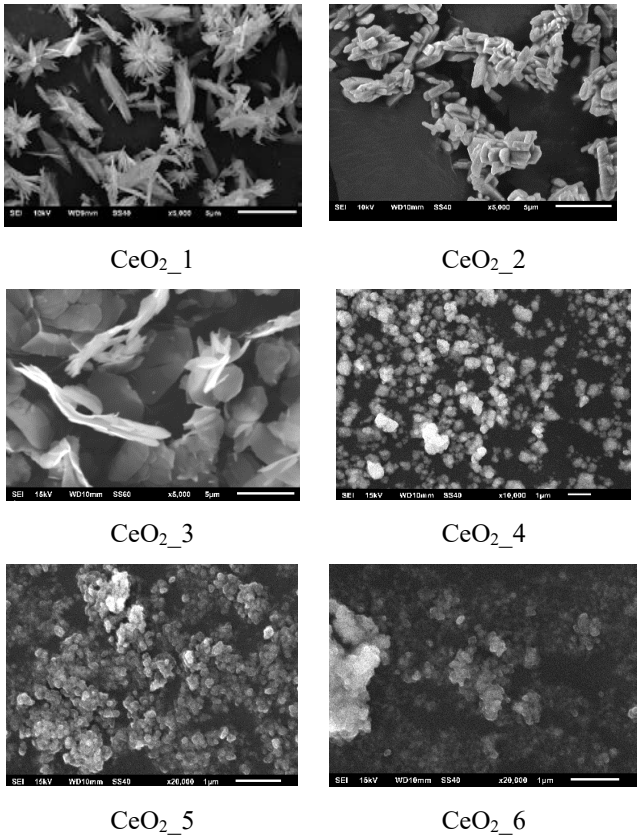


Figure 5. The SEM images of the CeO_2 samples under magnification of 5000x-20000x.

Figure 6. The Raman spectra of CeO_2_2 sample (crystalline) and CeO_2_6 sample (amorphous).

Raman spectroscopy was used to compare the structural ordering of selected CeO_2 samples. The crystalline CeO_2_2 sample exhibited a strong and sharp band at approximately 460–465 cm^{-1}, corresponding to the F2g mode of fluorite-type CeO_2, confirming the formation of a well-crystallized ceria phase. In contrast, the CeO_2_6 sample showed no dominant sharp F2g band, but rather a broadened and more complex spectral profile, indicating the absence of long-range crystalline order and confirming its amorphous or poorly crystalline character.

The pronounced spectral differences between these two samples demonstrate the strong influence of synthetic history and thermal treatment on the structural organization of CeO_2 nanoparticles.

Table 3. The specific surface area of the selected CeO_2 samples

	Sample		
	CeO2_3	**CeO2_5**	**CeO2_6**
Slope	30.8528	71.6628	69.0467
Intercept	0.274273	0.884697	1.12907
Correl. coeff.	0.999949	0.999857	0.999951
C constant	113.489	82.0026	62.1533
Surface area	111.881 m²/g	48.003 m²/g	49.626 m²/g

In the Table 3, the specific surface area obtained by N_2 adsorption isotherms fitted by BET method are given. Nitrogen adsorption measurements revealed pronounced differences in the specific surface area of the selected CeO_2 samples. The highest BET surface area was obtained for CeO_2_3, synthesized by precipitation with ammonium carbonate, reaching 111.9 $m^2\ g^{-1}$. This value is more than two times higher than those measured for CeO_2_5 and CeO_2_6, which exhibited specific surface areas of 48.0 $m^2\ g^{-1}$ and 49.6 $m^2\ g^{-1}$, respectively. The high correlation coefficients of the BET plots confirmed the reliability of the linear fitting in the selected pressure range. The markedly larger surface area of CeO_2_3 is consistent with its open flake-like morphology observed by SEM and may contribute to the higher availability of surface-active sites for HD degradation.

3.3. The HD Degradation experiments

The removal/degradation profiles of HD over the investigated CeO_2 samples are presented in Figure 7. They clearly demonstrate that the catalytic performance strongly depends on the synthetic route. Among all samples, CeO_2_3 showed the highest degradation efficiency, reducing the relative HD concentration to approximately 0.16 after 70 min. A pronounced decrease was already observed during the first 20 min, indicating rapid interaction of HD with the surface-active sites. This superior performance is in good agreement with the SEM observations, where CeO_2_3 exhibited open lamellar/flake-like morphology with exposed edges, and with BET analysis, which showed the highest specific surface area for this sample, 111.9 $m^2\ g^{-1}$. Such morphology and high surface area may provide a larger number of accessible adsorption and reaction sites for HD degradation. CeO_2_4, obtained by solution combustion synthesis, also exhibited relatively high degradation efficiency, reaching a final relative HD concentration of approximately 0.34. Its activity may be associated with the

fine granular morphology and porous agglomerate structure typically formed during combustion synthesis. These features may favor the formation of defect-rich surfaces and facilitate the accessibility of HD molecules to catalytically active sites.

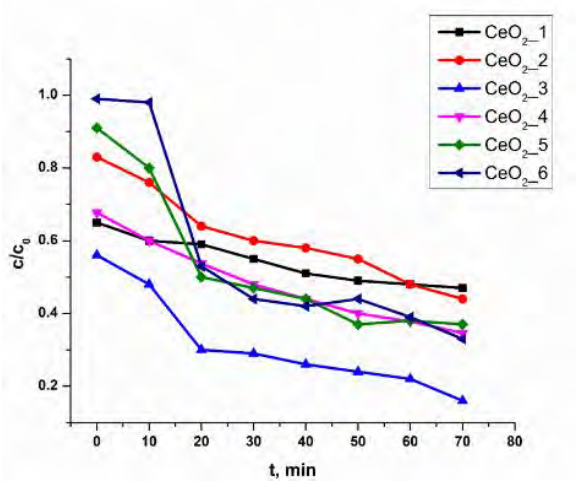


Figure 7. The HD removal over CeO_2 samples

In contrast, CeO_2_1 and CeO_2_2 showed lower degradation efficiency, despite their clearly developed morphologies. This suggests that anisotropic or crystalline morphology alone is not sufficient for high catalytic activity if the accessible surface area and density of reactive surface defects are not optimal. The comparison between CeO_2_2 and CeO_2_6 further indicates that crystallinity alone does not govern HD degradation. Raman spectroscopy confirmed that CeO_2_2 is crystalline, showing the characteristic fluorite CeO_2 band, while CeO_2_6 is amorphous or poorly crystalline. Nevertheless, CeO_2_6 showed considerable HD removal, especially after the first 20 min, despite its lower specific surface area of about 49.6 $m^2\ g^{-1}$ compared with CeO_2_3. This behavior may be related to the higher degree of structural disorder and defect concentration in amorphous CeO_2. However, its strongly agglomerated morphology probably limits the accessibility of these active sites, which may explain why it does not reach the performance of CeO_2_3. In order to obtain full degradation data with the posibility for the kinetic modelling and estimation which part of the HD removal includes adsorption and which is catalytic degradation, additional experiments should be performed with additional time points between 0 and 20 minutes.

4. CONCLUSONS

Overall, the results indicate that HD removal over CeO_2 nanoparticles is controlled by a combination of factors, including specific surface area, morphology, crystallinity, structural disorder, and probably the concentration of oxygen vacancies and Ce^{3+} surface sites. The high activity of CeO_2_3 appears to arise from the favorable combination of high surface area and open flake-like morphology, whereas the activity of CeO_2_4 may be related to its porous, combustion-derived, defect-rich structure. Therefore, the catalytic behavior cannot be explained solely by crystallinity or surface area, but rather by the accessibility and nature of surface-active sites.

Further research will encompass the incorporation of the most effective CeO_2 nanoparticles in the electrospun nanofibers of various polymers, as well as their application onto various substrates intended for CBRN protection, so the further work will consider the possibility of the scale-up of their application and use.

Acknowledgements

This work was supported by the Ministry of Science, Technological development and Innovations of the Republic of Serbia, Contract No. 451-03-34/2026-03/200325.

References

[1] E. M. Lucas and K. J. Klabunde, "Nanocrystals as destructive absorbants for mimcs of chemical warfare agents," *Nanostructured Materials*, vol. 12, no. 1, pp. 179–182, 1999, doi: https://doi.org/10.1016/S0965-9773(99)00093-8.

[2] A. Khaleel, W. Li, and K. J. Klabunde, "Nanocrystals as stoichiometric reagents with unique surface chemistry. New adsorbents for air purification," *Nanostructured Materials*, vol. 12, no. 1, pp. 463–466, 1999, doi: https://doi.org/10.1016/S0965-9773(99)00159-2.

[3] V. Štengl, T. M. Grygar, F. Opluštil, and T. Němec, "Ge4+ doped TiO2 for stoichiometric degradation of warfare agents," *J. Hazard. Mater.*, vol. 227–228, pp. 62–67, 2012, doi: https://doi.org/10.1016/j.jhazmat.2012.05.007.

[4] V. Štengl *et al.*, "Zirconium doped nano-dispersed oxides of Fe, Al and Zn for destruction of warfare agents," *Mater. Charact.*, vol. 61, no. 11, pp. 1080–1088, 2010, doi: https://doi.org/10.1016/j.matchar.2010.06.021.

[5] K. Ghabili, P. S. Agutter, M. Ghanei, K. Ansarin, Y. Panahi, and M. M. Shoja, "Sulfur mustard toxicity: History, chemistry, pharmacokinetics, and pharmacodynamics," *Crit. Rev. Toxicol.*, vol. 41, no. 5, pp. 384–403, May 2011, doi: 10.3109/10408444.2010.541224.

[6] Gary S Groenewold, "Degradation kinetics of VX," *Main Group Chemistry*, vol. 9, no. 3–4, pp. 221–244, Nov. 2010, doi: 10.3233/MGC-2010-0037.

[7] K. Otsuka, T. Ushiyama, and I. Yamanaka, "Partial Oxidation of Methane Using the Redox of Cerium Oxide," *Chem. Lett.*, vol. 22, no. 9, pp. 1517–1520, Sep. 1993, doi: 10.1246/cl.1993.1517.

[8] S. Park, J. M. Vohs, and R. J. Gorte, "Direct oxidation of hydrocarbons in a solid-oxide fuel cell," *Nature*, vol. 404, no. 6775, pp. 265–267, 2000, doi: 10.1038/35005040.

[9] V. Baldim, F. Bedioui, N. Mignet, I. Margaill, and J.-F. Berret, "The enzyme-like catalytic activity of cerium oxide nanoparticles and its dependency on Ce3+ surface area concentration," *Nanoscale*, vol. 10, no. 15, pp. 6971–6980, Apr. 2018, doi: 10.1039/c8nr00325d.

[10] Y. Yang *et al.*, "Redox enzyme-mimicking activities of CeO2 nanostructures: Intrinsic influence of exposed facets," *Sci. Rep.*, vol. 6, no. 1, p. 35344, 2016, doi: 10.1038/srep35344.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF AN INCREASED HOPCALITE LAYER CONTENT IN THE ACTIVE FILLING ON THE SORPTION CHARACTERISTICS OF THE FILTER

BILJANA MIHAJLOVIĆ
TRAYAL Corporation, Kruševac, aktivni.ugljevi@gmail.com

TATJANA MARKOVIĆ
Military technical institute, Belgrade, tanjin.mejl@gmail.com

MARINA ILIĆ
Military technical institute, Belgrade, marinailic970@gmail.com

VUKICA GRKOVIĆ
TRAYAL Corporation, Kruševac, vukica.g30@gmail.com

VLADA PETROVIC
TRAYAL Corporation, Kruševac, fzstrayal@gmail.com

Abstract: *In recent years, an alternative CO CBRN ESC filter has been developed, differing from the standard military filter in the composition of its active bed by incorporating an additional hopcalite layer. Testing of its sorption characteristics revealed that the CO CBRN ESC filter did not achieve the required protection times against certain test gases. To improve the sorption performance, a new CO CBRN filter was developed. Compared with the previous CO CBRN ESC filter, the volumetric ratio of the active bed layers was modified by increasing the proportion of the hopcalite layer. The sorption characteristics of the new CO CBRN filter were evaluated, and the results confirmed the effectiveness of this approach. Significantly longer protection times were achieved for all tested gases compared with those obtained using the previous filter design. Although the increase in hopcalite content resulted in a higher filter mass and consequently reduced some aspects of its functionality, the new filter remains suitable for integration into military protective masks. It can be employed both for military applications and for protection in emergency and accidental hazardous situations.*

Keywords: *CO CBRN filter development, two-component active bed, increased hopcalite layer, sorption characteristics, functional characteristics.*

1. INTRODUCTION

Respiratory protection is a complex and multidisciplinary field that requires a systematic approach to the selection of suitable filters capable of ensuring adequate protection of the respiratory system under various hazardous conditions. The proper choice of a filter is of critical importance in emergency situations, industrial accidents, as well as in potential terrorist incidents involving toxic chemical agents. Current global safety and security challenges have emphasized the need for the development of alternative filter designs with broadened protective capabilities. Such filters should provide effective protection against a wide spectrum of toxic chemicals, while maintaining satisfactory performance characteristics.

In a previous study, TRAYAL Company, in collaboration with Military technical Institute, developed an alternative combined filter designated as CO CBRN ESC. This filter differs in the composition of its active bed compared to the standard military M7 filter and conventional CBRN filters. The active bed of the CO CBRN ESC filter consists of two distinct components: a layer of activated carbon impregnated with calcium chloride and a hopcalite layer, arranged in a volumetric ratio of 2:1 with a standard bed height.[1]

To simulate conditions representative of accidental exposure scenarios, the gas flow rate was varied during experimental testing, while parameters such as relative humidity of the gas mixture and inlet test gas concentration were kept constant throughout the experiments. The results obtained showed that the alternative filter did not meet the required 15-minute protection time at higher gas flow rates, in accordance with the NIOSH (National Institute for Occupational Safety and Health, USA) requirements for escape respirator filters. This was observed for the following test gases: phosgene, hydrogen cyanide, sulfur dioxide, and carbon monoxide. Therefore, the filter cannot be used as an escape-type respirator.

Corresponding author: T. Marković;
E-mail: tanjin.mejl@gmail.com

DOI: 10.5937/OTEH26130M

This raises several key questions regarding the required filter characteristics necessary to ensure extended protection time at a constant gas concentration under varying gas flow rates. In particular, it is important to determine which flow rates provide more realistic information on user protection during different exposure scenarios, such as normal walking or running conditions, as well as what filter mass can be considered acceptable in order to achieve improved protection without significantly compromising functional performance.[2]

Based on the results obtained in the previous study of the CO CBRN ESC filter, it was decided not to introduce major or costly modifications to the design of the new alternative combined filter, designated CO CBRN.[3]

Compared to the previous CO CBRN ESC filter, improvements in the sorption characteristics of the CO CBRN filter were achieved by modifying the volumetric ratio of the active bed layers. The volume of the activated carbon impregnated layer was kept unchanged, while the hopcalite layer was increased to its maximum packing height (approximately corresponding to the CBRN filter standard height). Consequently, the volume of the hopcalite layer in the new CO CBRN filter was doubled compared to that in the previous design.protection time at higher gas flow rates, in accordance with the NIOSH (National Institute for Occupational Safety and Health, USA) requirements for escape respirator filters. This was observed for the following test gases: phosgene, hydrogen cyanide, sulfur dioxide, and carbon monoxide. Therefore, the filter cannot be used as an escape-type respirator.

By increasing the hopcalite layer in the CO CBRN filter, as well as its total mass, an improvement in protection time across a broader range of test gases can be reasonably expected. However, such modifications are also associated with a reduction in certain functional characteristics during use. Despite the reduced operational functionality of the CO CBRN filter, it meets the mass requirement of up to 500 g for full-face masks according to the relevant standard SRPS EN 14387.

The testing of the new CO CBRN filter was conducted under the same defined and controlled experimental conditions as those applied to the previous filter. The experimental protocol included a constant inlet concentration, variation of test gas flow rates, as well as controlled humidity and temperature of the air–test gas mixture, which partially simulates conditions encountered in accidental exposure scenarios.

2. MATERIALS AND METHODS

The activated carbon is produced using TRAYAL Company technology, with a particle size distribution corresponding to the M3 filter and an iodine number not lower than 1250 mg/g. Hopcalite, as the second component, consists of a mixture of copper oxide (CuO) and manganese dioxide (MnO_2) and is defined by its physical and chemical characteristics, while its homologation is carried out within the manufacturing facility.[4]

The alternative CO CBRN combined filters used for sorption testing against selected test gases were filled with the described two-component active bed and manufactured using standard M3 filter production technology. In addition to the modification of the active bed layer (increased hopcalite volume), a modified anti-aerosol insert was introduced to enable more efficient utilization of the available space for the active filling. [5] The sorption performance of the newly developed co cbrn filter with respect to selected test gases was investigated in compliance with the methods specified by standards SORS 8748 and SRPS EN 14387.[6][7]

The testing methods involve monitoring and recording the protection time as an indicator of the sorption characteristics of the filter during exposure to test gases under defined experimental conditions and parameters. These methods include maintaining a constant inlet concentration of test gases and a controlled relative humidity of the gas–air mixture, while varying the flow rate of the inlet gas stream.

3. RESULTS AND DISCUSSION

The developed alternative combined CO CBRN filter contains a two-component active bed consisting of a layer of activated carbon impregnated with calcium salts ($CaCl_2$) and a hopcalite layer, arranged in a volumetric ratio of 1:1. Hopcalite is a mixture of copper and manganese oxides (Cu^{2+} and Mn^{4+}).

The activated carbon used in the filter is produced using a technology developed by TRAYAL Company, while the hopcalite of defined quality is supplied by a long-term external provider. As part of the standard quality control procedure for incoming materials, the hopcalite is tested in accordance with the relevant technical specification, and its physicochemical properties are verified immediately prior to filter filling.

3.1. Results of the Physical and Mechanical Characterization of Both Active Bed Components and the Filter

In order to obtain a clearer understanding of the potential broader application of the CO CBRN filter in different protection domains, as well as the possibility of improving its functional performance, it is necessary to evaluate the physical and mechanical properties of the materials of both active bed layers (activated carbon and hopcalite) prior to filter filling, in accordance with the SORS 8830/05 standard.[8][9] The results of the physical and mechanical characterization of the active bed materials used in the new CO CBRN filter are presented in Table 1.

Table 1. Physical and Mechanical Properties of CO CBRN Active Bed Materials

Physical and Mechanical Properties of Material	Impregnated activated carbon	Hopcalite
Moisture content (%)	1	0,04
Bulk density (g/L)	485	803

Mechanical strength (%)	84	63
Particle size distribution (%)		
>1.8 mm	0.00%	9,71%
1.6-1.8 mm	7,53%	20.19%
1.4-1.6 mm	20,86.%	24.04%
0.7-1.4 mm	71,50%	45.81%
0.5-0.7 mm	0.02%	014%
<0.5 mm	0.09%	0.11%

Considering that both activated carbon and hopcalite are granular materials in structure, and based on the results presented in Table 1, the physical and mechanical properties of the active bed materials of the new CO CBRN filter are comparable to those of the active bed materials used in the CO CBRN ESC filter. Compared to the previous CO CBRN ESC filter, the new CO CBRN filter retains the same volume of the impregnated activated carbon layer, while the volume of the hopcalite layer has been increased.

The decision to increase the proportion of the hopcalite layer in the new filter was based on the inferior protection time performance observed for the CO CBRN ESC filter against certain test gases. In the CO CBRN filter, the hopcalite layer was increased to achieve an approximately equal contribution of both active bed layers, while the total volume of the active bed was adjusted to match the existing filter housing design up to its maximum filling height. After the evaluation of the physical and mechanical properties of the active bed materials and assembly of the new CO CBRN filter, the filter was further tested for total mass, inhalation resistance, and filtration efficiency against paraffin oil mist.

Table 2. Functional Characteristics of the CO CBRN Filter

Filters COCBRN	Mass (g)	Inhalation resistance (Pa) 30 L/min i 95 L/min	Filtration efficiency (%)
Sample 1	385	188/720	$2x10^{-3}$
Sample 2	397	185/764	$1x10^{-3}$
Sample 3	404	190/850	$1.5x10^{-3}$

The results presented in Table 2 indicate the following characteristics of the newly developed combined CO CBRN filter:

• Compared to the previous CO CBRN ESC filter, the first observed change is an increase in the total filter mass of approximately 80 g, primarily due to the increased hopcalite layer. As a consequence, the CO CBRN filter is not suitable for use in half-mask respirators or children's masks.

• The changes in the height and composition of the active bed also lead to a significant change in the inhalation resistance of the filter.

• The obtained values of filtration efficiency against paraffin oil mist remained unchanged, since the anti-aerosol layer was retained from the previous CO CBRN ESC filter design.

The presented results indicate that the newly developed combined CO CBRN filter complies with the requirements of the SRPS EN 14387 standard in terms of total mass (up to 500 g for full-face masks) and inhalation resistance (2P3). Accordingly, further testing of the sorption characteristics of the filter against standard-defined test gases can be continued.

These findings confirm that the alternative active bed configuration can be applied in filter manufacturing, thereby providing a realistic potential for practical implementation, particularly in emergency situations where user response and operational conditions may vary.

3.2. Results of the Sorption Characteristics Testing of the CO CBRN Filter against NBC Test Gases (Phosgene, Chloropicrin, and Hydrogen Cyanide)

The testing of the sorption characteristics of the newly developed combined CO CBRN filter against NBC test gases (phosgene, hydrogen cyanide, and chloropicrin), expressed as protection time (in minutes), was conducted under the following conditions:

• inlet concentration of test gases (phosgene and hydrogen cyanide): 1000 ppm,
• inlet concentration of test gas (chloropicrin): 2200 ppm,
• relative humidity of the air–gas mixture (phosgene, hydrogen cyanide, and chloropicrin): 80%, and
• air–gas mixture flow rates: 30 L/min and 64 L/min.

The inlet concentrations of the test gases, as well as the relative humidity of the incoming air–gas mixture, were kept constant throughout all filter samples tested. In order to simulate conditions closer to accidental exposure scenarios, the sorption characteristics were evaluated at different gas flow rates of 30 L/min and 64 L/min.

The obtained results of the sorption performance of the CO CBRN filter against NBC test gases (phosgene, hydrogen cyanide, and chloropicrin), expressed as protection time (in minutes), are presented in Table 3 (relative humidity of the gas–air mixture: 80%).

Table 2. Functional Characteristics of the CO CBRN Filter

Test -gas	Inlet concentration of the gas mixture (ppm)	Protection time (min) 30 L/min	Protection time 64 L/min
Phosgene	1000	45 46 48	18 19 20
Chloropicrin	2200	38 38 39	29 29 30
Hydrogen cyanide	1000	>50 >53 >55	25 27 26

By comparing the protection times of the CO CBRN filter against the three test gases with those obtained for the previous CO CBRN ESC filter, it can be observed that the

increase in the hopcalite layer (by nearly twofold) leads to a significant improvement in the protection time of the CO CBRN filter, in some cases even resulting in a doubling of performance, particularly for phosgene and hydrogen cyanide.

For both the CO CBRN filter and the previous CO CBRN ESC filter, Figures 1 and 2 illustrate the difference in protection times against phosgene at different flow rates (30 L/min and 60 L/min).

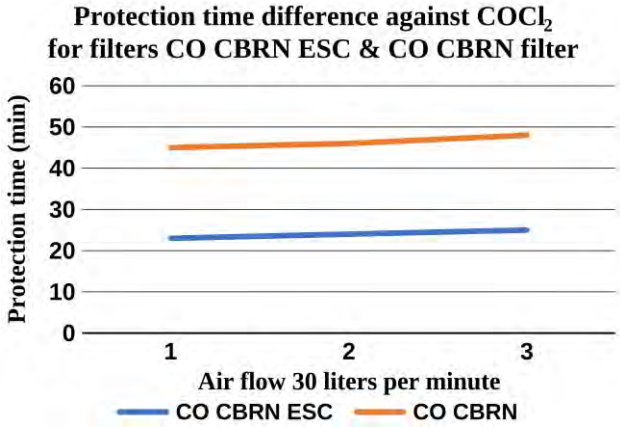


Figure 1. Difference in time protection against phosgene on air flow 30 L/min

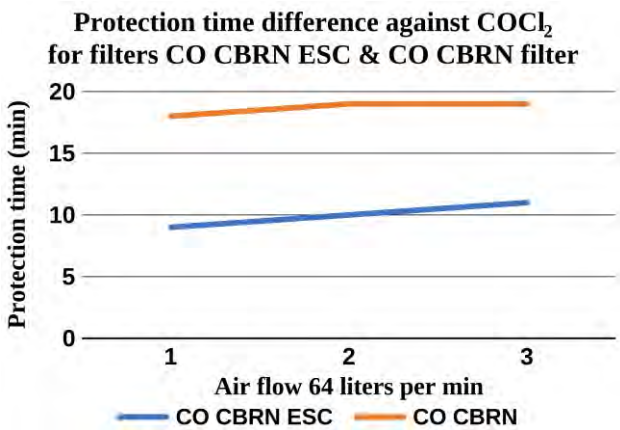


Figure 2. Difference in protection time against phosgene under an air flow rate of 64 L/min

3.3. Results of the Sorption Performance Evaluation of the CO CBRN Filter against Industrial Toxic Gases (Sulfur Dioxide, Cyclohexane, Ammonia, and Carbon Monoxide)

The sorption characteristics of the combined CO CBRN filter against industrial toxic gases (sulfur dioxide SO_2, cyclohexane C_6H_{12}, and ammonia NH_3), expressed in terms of protection time, were evaluated under the following test conditions:

- Inlet concentration of the test gases: 1000 ppm
- Relative humidity of the air-gas mixture: 70% (until breakthrough)
- Flow rate of the air–gas mixture: 30 L/min and 64 L/min

The sorption characteristics of the combined CO CBRN filter against industrial toxic gas carbon monoxide CO expressed in terms of protection time, were evaluated under the following test conditions:

- Inlet concentration of the test gas: 2500 ppm
- Relative humidity of the air-gas mixture: 80% (until breakthrough at 200 ppm and temperature 50 °C)
- Flow rate of the air–gas mixture: 30 L/min and 45 L/min

Due to the limitations of the test apparatus used for evaluating the sorption characteristics of the filter against carbon monoxide, a flow rate of 64 L/min could not be achieved. The system operates using an artificial lung at 20 cycles per minute and 1.5 L per cycle, while the maximum achievable setting is 30 cycles per minute, corresponding to 45 L/min. The obtained results of the sorption performance of the CO CBRN filter against industrial test gases, expressed as protection time, are presented in Table 4.

Comparison of the protection times of the CO CBRN filter against the four industrial test gases with those obtained for the previous CO CBRN ESC filter shows that the increase in the hopcalite layer leads to a significant improvement in protection performance, particularly for sulfur dioxide and carbon monoxide.

For both the new CO CBRN filter and the previous CO CBRN ESC filter, Figures 1 and 2 illustrate the difference in protection times against sulfur dioxide at different flow rates (30 L/min and 60 L/min). The observed difference in protection time for sulfur dioxide is approximately four times greater at the lower flow rate of 30 L/min, while at the higher flow rate the difference is less pronounced.

Table 4. Sorption Characteristics of the CO CBRN Filter against Industrial Test Gases

Test-gas	Relative humidity of the air-gas mixture (%)	Inlet concentration of the test gas (ppm)	Protection time (min) 30 L/min	Protection time (min) 64/ 45* l/min
SO_2	70	1000	126 128 126	25 26 26
NH_3	70	1000	95 100 97	48 50 46
C_6H_{12}	70	1000	78 75 85	34.3 29.1 33
CO	80	2500	50 54 48	60 59 59

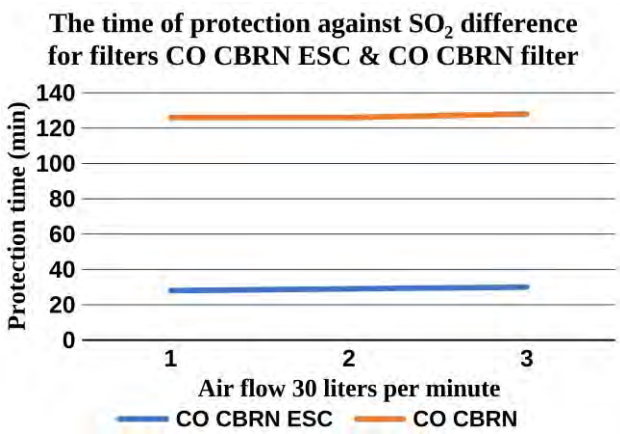


Figure 3. Difference in time protection against SO_2 on air flow 30 L/min

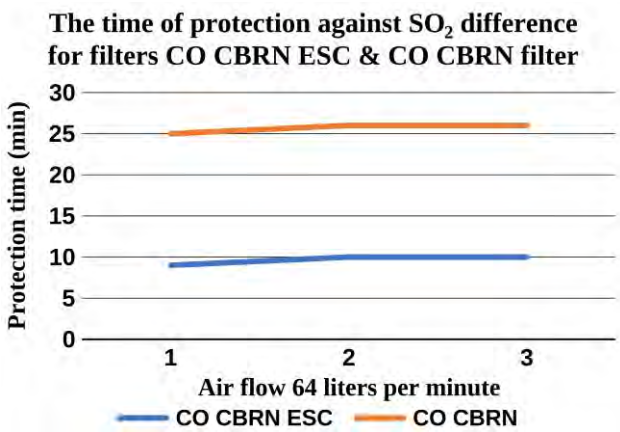


Figure 4. Difference in protection time against SO_2 under an air flow rate of 64 L/min

Given that the tests have shown that, in accordance with the American standard for escape respirator filters, the CO CBRN filter achieves the required 15-minute protection time at higher gas flow rates and under the specified inlet concentrations, it can be concluded that the filter, when integrated into a full-face respirator, has potential applications both in accidental scenarios and fire-related emergencies, as well as in military use.

Preliminary tests of the protection time of the CO CBRN filter against the above-mentioned test gases were also conducted under conditions of increased inlet gas mixture concentrations (5000 ppm) and a lower flow rate of 30 L/min. The obtained results are presented in Table 5. The protection time results indicate the potential applicability of the CO CBRN filter in environments where toxic gas concentrations are significantly higher than the initially defined test conditions (1000 ppm and 2500 ppm).

Table 5. Sorption Characteristics of the CO CBRN Filter against Industrial Test Gases at 5000 ppm

Test gas	Relative humidity [%]	Protection time [min] 30 L/min
SO_2	70	29 28
NH_3	70	26 22
C_6H_{12}	70	38 39
CO/	80	30 32

4. CONCLUSION

In a previous study, the results of the sorption characteristics testing of the CO CBRN ESC filter showed that it did not meet the required 15-minute protection time at higher gas flow rates for the following test gases: phosgene, hydrogen cyanide, sulfur dioxide, and carbon monoxide. Consequently, the filter cannot be used as part of a military respirator for emergency situations in which fire-related hazards may also occur.

In order to improve the protection time of the filter against the above-mentioned test gases, a new CO CBRN filter was developed. Compared to the previous CO CBRN ESC filter, the ratio of the active bed layers was modified in favor of the hopcalite layer, with the bed height adjusted to the maximum filling height corresponding to a standard CBRN filter design. The sorption characteristics of the CO CBRN filter were evaluated against NBC agents and toxic industrial gases under the same experimental conditions as those applied for the CO CBRN ESC filter.

In addition, the obtained results for the CO CBRN filter suggest that, due to the basic nature of hopcalite, an increase in protection time can also be expected for other acidic test gases such as chlorine (Cl_2) and hydrogen sulfide (H_2S).
The increased capacity and bed height of the new CO CBRN filter, resulting from the higher hopcalite content, also indicate the potential for adequate protection under exposure to higher gas concentrations (up to 5000 ppm).

However, the increase in the hopcalite layer in the CO CBRN filter also leads to an increase in its total mass, which results in reduced functional performance during use. Despite this limitation, the filter can still be used as an integral component of full-face respirators for both military applications and emergency situations in peacetime. Due to the increased total mass, the CO CBRN filter is not suitable for use in half-mask respirators, children's masks, or evacuation hoods.

References

[1] MIHAJLOVIC,B.,ILIC,M.,MARKOVIC,T., GRKOVIC,V.: Impact of active filler composition modification on the sorption characteristics of filters, OTEH Belgrade, 2024.

[2] ILIC,M.,SENIC,Z.,PETROVIC,V., MIHAJLOVIC,B,GRKOVIC,V: Impact of the altered texture of the active filling of the filter on the sorptive characteristics with the special reference to the efficiency of filtering, OTEH, Belgrade, 2016.

[3] ILIC,M.,MARKOVIC,T.,MIHAJLOVIC,B, GRKOVIC,V, BAUK,S.: *Impact of the different active filling height of the filter on the sportive characteristics of filtering*, OTEH Tara, 2022..

[4] SMISEK,M., Cerny,S., adopted from Active Carbon, Elsevier Publ. Co., Amsterdam, 1970. and JANKOWSKA,H., SWIATKOWSKI,A., CHOMA, J.: *Active Carbon, Ellis Howard, England, 1991.*

[5] TAYLOR,E.: *Letal Mists, An Introduction to the natural and Militar Sciences of Chemical, Biological Warfare and Terrorism*

[6] SORS 8748:2004, *Personal CBRN protect device Requirements and methods testing*

[7] SRPS EN 14387:2013, Respiratory protective *devices- Gas filter(s) and combined filter(s)*

[8] SORS 8829:2005, Filter, Combined filter M3, *Requirements and methods testing*

[9] SORS 8830:2005,*Personal CBRN protect device Active carbon KI M3 for combined filter, Requirements and methods testing*.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INFLUENCE OF PROCESS PARAMETERS ON QUALITY OF LASER POWDER BED FUSION PRODUCED AL10SIMG PARTS

ALEKSANDAR ĆITIĆ
Military Technical Institute, Belgrade, aleksandar.citic@mod.gov.rs

VESNA LUKIĆ
Military Technical Institute, Belgrade

JOVANA MANDIĆ
Military Technical Institute, Belgrade

MARKO BOJIN
Military Technical Institute, Belgrade

ALEKSA GRUBIĆ
Military Technical Institute, Belgrade

MILOŠ STEPANOVIĆ
Military Technical Institute, Belgrade

ŽIVORAD JEVTIĆ
Military Technical Institute, Belgrade

Abstract: *This study examines the influence of Laser Powder Bed Fusion (LPBF) process parameters on the quality of AlSi10Mg alloy parts. Two sets of samples were fabricated using different combinations of laser power, scan speed, and hatch spacing, resulting in distinct energy densities. Visual examination revealed that lower energy density resulted in rough surfaces and delamination, indicating insufficient interlayer bonding. Chemical analysis confirmed that all samples met EN AC-Al10SiMg compositional requirements, while hardness measurements showed no significant variation between parameter sets or build orientations. Microstructural observations identified typical LPBF features, including fine cellular Al–Si eutectic and coarse dendritic like grains, with pores more frequent in low-energy samples. The results demonstrate that LPBF process parameters strongly affect surface morphology while having negligible impact on bulk hardness. Optimizing energy density is essential for achieving improved microstructural uniformity and overall part quality in AlSi10Mg components.*

Keywords: *LPBF, Al10SiMg, Process parameters*

1. INTRODUCTION

Additive Manufacturing (AM) enables the production of complex geometries directly from a digital model, without the need for traditional tools or molds. Among various AM techniques, Laser Powder Bed Fusion (LPBF) is most adopted method for manufacturing metal components. The LPBF process uses a high energy laser beam to selectively melt a layer of metal powder, forming parts layer by layer [1, 2] under extremely high cooling rates on the order of $10^6 - 10^8\ ^\circ C/s$. Such thermal conditions lead to the formation of a unique microstructure, characteristic of LPBF process, which differs significantly from the microstructures of alloys produced by conventional processes such as casting, rolling or forging [3]. Wide range of metal powders for use in LPBF process is developed such as 316L steel, maraging steel, titanium, nickel, aluminum alloys and others. Al alloys offer good compromise between strength and density, in addition to being inexpensive. Most common Al alloy used in LPBF is Al10SiMg, a hypoeutectic aluminum alloy containing approximately 10% Si with an addition of Mg. This alloy is recognized as highly suitable for AM due to its excellent processability, low density, good corrosion resistance, and favorable mechanical properties. Compared with other metal powders, aluminum alloys still present challenge in the LPBF process due to high reflectivity, high thermal conductivity and low melt viscosity [1, 2, 4]. In the LPBF process, AlSi10Mg forms a fine cellular Al–Si eutectic, with a silicon network distributed around primary aluminum cells, resulting in increased strength and hardness compared to cast alloy states. However, this microstructure simultaneously leads to reduced ductility and increased sensitivity to residual stresses [4]. Components produced from AlSi10Mg are widely used in automotive engineering, aerospace structures, robotics, energy systems, and lightweight mechanical assemblies, where the combination of low weight and high strength is essential [5]. Despite these advantages, the final quality of

Corresponding author: A. Ćitić;
E-mail: aleksandar.citic@mod.gov.rs
DOI: 10.5937/OTEH26131C

LPBF-fabricated AlSi10Mg parts is sensitive to process parameters, which govern melt pool behavior, microstructural evolution, defect formation, and mechanical properties. Key LPBF parameters include laser power, scan speed, hatch spacing, layer thickness, preheating temperature, and scanning strategy. Together, these parameters determine the volumetric energy density delivered to the powder bed, influencing melt pool stability, track geometry, layer bonding, and porosity formation. Insufficient energy input typically leads to lack-of-fusion, while excessive energy can cause keyhole porosity, evaporation of alloying elements, and microstructural inhomogeneity [6]. Hatch spacing affects the overlap between adjacent melt tracks, while scanning strategy influences residual stress distribution and grain texture. Therefore, optimizing process parameters is essential for achieving high density, uniform microstructure, and reliable mechanical properties. In addition to process parameters, the quality of AlSi10Mg powder plays a crucial role. Powder used for LPBF must exhibit spherical morphology, narrow particle size distribution, low oxide content, and good flowability, as deviations can lead to segregation, non-uniform layer deposition, and increased porosity [7]. Recycled powder, although economically attractive, often shows increased oxygen content and altered particle size distribution, which may degrade the mechanical properties of printed parts [8]. Considering all the above, LPBF process parameters have a decisive influence on the quality of AlSi10Mg parts, including their density, microstructure, mechanical properties, defect formation, and long-term reliability. Understanding the interdependence of these parameters is fundamental for process optimization and improving the characteristics of final components.

2. EXPERIMENTAL PROCEDURE

2.1 Material and LPBF process

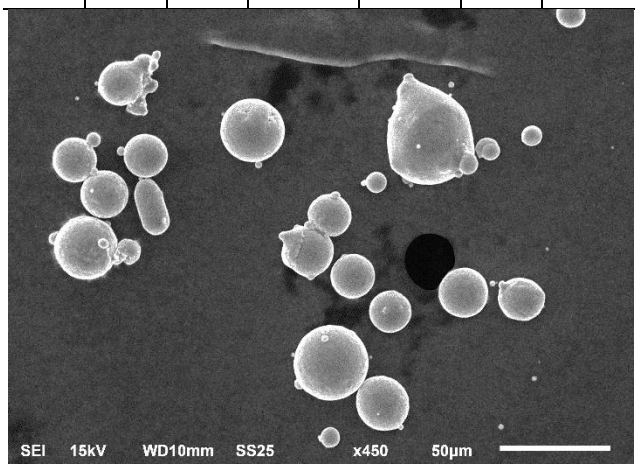
To produce additively manufactured samples, Eplus 3D EP-M450 machine was used. Two sets of samples made of Al10SiMg powder were produced with a different set of process parameters, (Table 1). The laser energy density (E) related to the LPBF process is calculated according to formula [1]:

$$E = \frac{P}{V \cdot H \cdot L_T} \qquad (1)$$

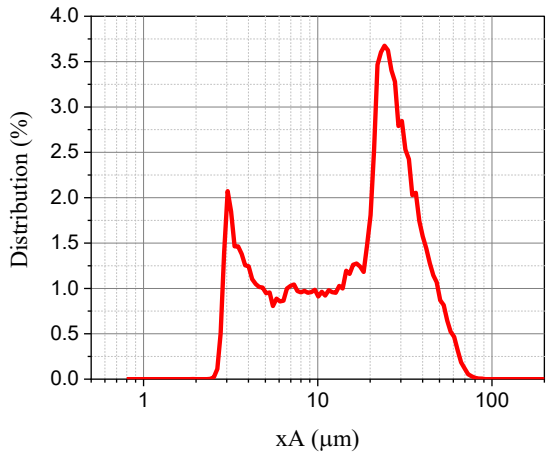
where P is laser power, V is the laser scan speed, H represents hatch distance and L_T is layer thickness. For both sets of sample's, layer thickness was held constant, while other parameters were changed, which gave energy density of 39.3 J/mm^3 for set I, and 50.3 J/mm^3 for set II parameters. For set I, nine identical cubes with dimensions (10x10x10) mm were produced (marked 1-9), and for set II, five identical cubes with dimensions (15x15x10) mm were produced (marked A-E). The morphology and particle size distribution of used powder is shown in Figure 1. Most of the observed powder particles are spherical, and the rest shows ellipsoidal or uneven shape. Particle size ranges from 3 µm to 70 µm, with mean size of 20.2 µm. 10 % of particles have size to 3,6 µm, 50 % of particles have size up to 20,3 µm while 90 % of particles are up to 38,9 µm. Mean value of particle circularity is 0,97. A JEOL JSM-6610LV scanning electron microscope was used to observe the morphology of used powder, while size and circularity were determined on Anton Paar Litesizer DIA 500 instrument

Table 1. Used process parameters

	Laser power P, W	Scan speed V, mm/s	Layer thickness L_T, µm	Hatch distance H, µm	Laser spot size, µm	Energy density E, J/mm^3
Set I	385	1400	50	140	85	39.3
Set II	425	1300	50	130	71	50.3



a) SEM image of powder particles



b) particle size distribution

Figure 1. Morphology and powder particle size distribution.

2.2 Characterization

Visual examination of the manufactured parts was performed after the samples were removed from the build plate, by naked eye and by Leica stereo microscope equipped with DFC 245 camera.

To determine if change in process parameters affected elements evaporation from the powder during melting, chemical composition of the produced samples was analyzed by Optical Energy Spectrometer (OES) Belec Compact Port.

Hardnes was determined by Brinell method (HB 2.5/62.5), according to SRPS EN ISO 6506 standard. Measurements were performed on XY plane (plane parallel to the build plate), and on XZ/YZ plane (plane parallel to the build direction).

For microstructure analysis samples were wet grounded on SiC paper, polished using diamond suspension and etched using Keller A reagent (1 ml HF, 1.5 ml HCL, 2.5 ml HNO_3 and 95 ml H_2O) for 8 s. Samples were observed in polished and etched state on Leitz Metaloplane optical microscope in bright field. Microstructure was analyzed in XY and XZ/YZ plane.

3. RESULTS

3.1. Visual examination

The surface morphology of produced samples is shown on Figure 2 for set I parameters and on Figure 3 for set II parameters. The surface of the samples produced using the parameter set I was visibly rougher than the surface of samples made with parameter set II. Observation under stereo microscope showed that the surface of parameter set I specimens was less uniform, with presence of coarse spherical particles (marked with red arrows) and fine powder particles (marked with yellow arrows), Figure 2a. Additionally, delamination of layers was observed on samples 7, 8, and 9, figure 2b-d.

Samples produced with Set II parameters showed smoother and more uniform surfaces. Although fine powder particles (marked with yellow arrows) were still present, Figure 3, no delamination was observed.

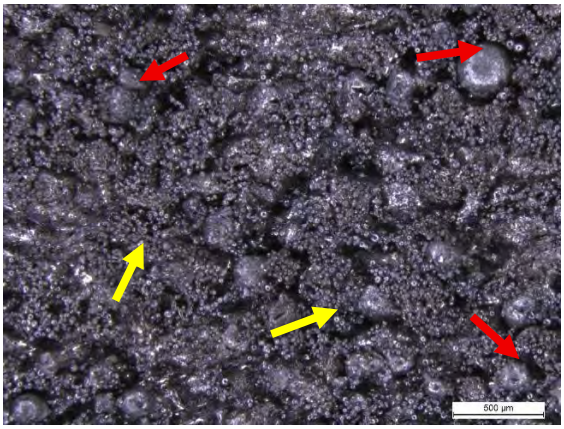
3.2 Chemical composition

The chemical composition of produced samples is given in Table 2.

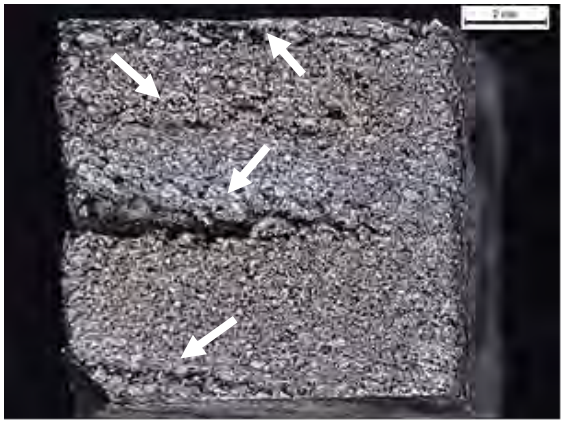
Table 2. Chemical composition of produced samples, mass. %

Element	Set I parameter samples	Set II parameter samples	EN AC-Al10SiMg according to SRPS EN 1706
Si	10.40	10.81	9.0-11.0
Fe	0.21	0.24	max. 0.55
Cu	<0.001	0.002	max. 0.05
Mn	0.008	0.007	max. 0.45
Mg	0.33	0.34	0.20-0.45
Ni	0.03	0.03	max. 0.05
Zn	<0,001	0.01	max. 0.10
Al	Balance	Balance	Balance

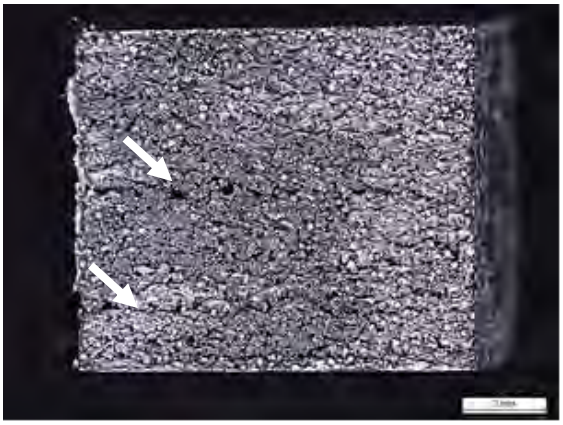
Chemical composition of samples from both sets of parameters is in accordance with standard SRPS EN 1706 for aluminum alloy EN AC Al10SiMg. No significant difference in chemical composition between samples from both sets of parameters is observed.



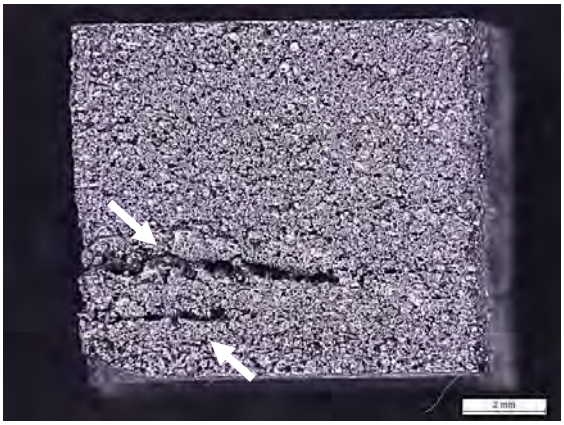
a) Sample No. 6



b) Sample No. 7

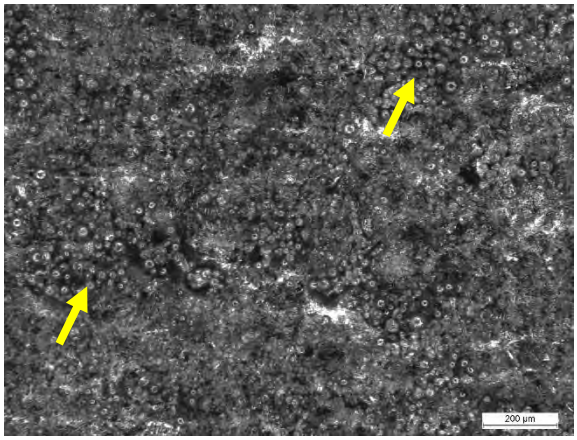


c) Sample No. 8

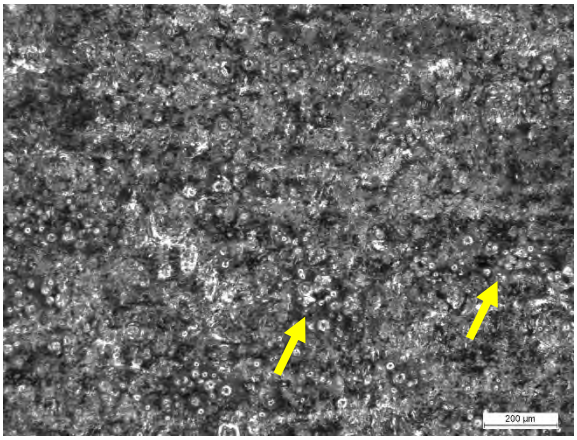


d) Sample No. 9

Figure 2. Surface of the samples produced with Set I parameters.



a) Sample A



b) Sample C

Figure 3. Surface of the samples produced with Set II parameters.

3.3 Hardness measurements

Results of hardness measurements are shown in table 3.

Table 3. Obtained hardness values, HB

	Measurement plane	Hardness, HB	Mean value
Set I	XY	129; 131; 129; 129; 131	130
	XZ/YZ	128; 128; 126; 128, 125	127
Set II	XY	130; 129; 130; 129; 133	130
	XZ/YZ	119; 119; 120; 119	119

3.4 Microstructure

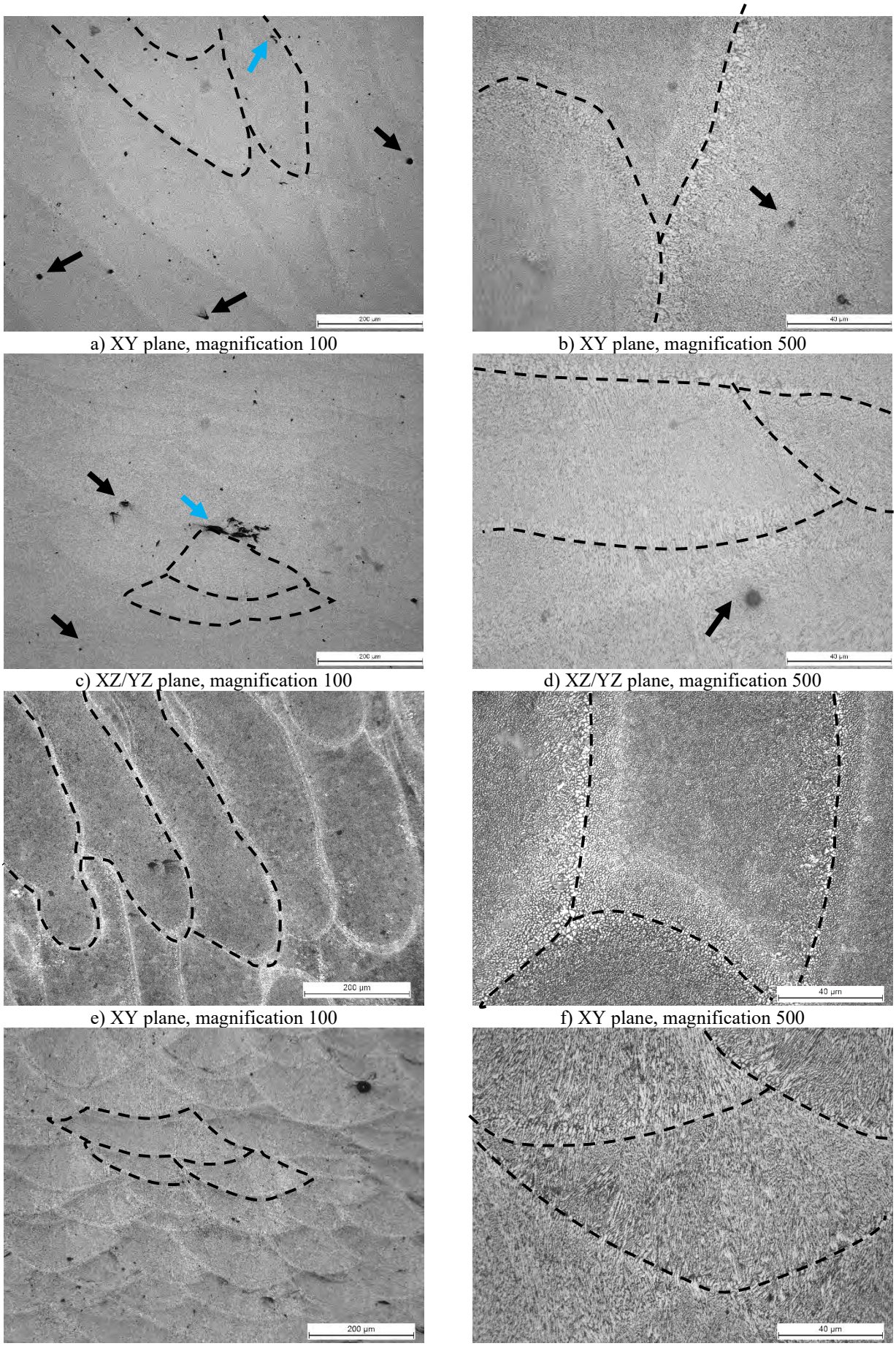
Figure 4 shows typical microstructure of specimens produced using set I parameters in as polished and etched state. Metallurgical defects such as gas pores (marked with black arrows) and lack of fusion (marked with blue arrows) were observed. Apart from the defects, clear melt pool boundaries were also observed in as polished state as well as after etching. On Figure 5 is shown microstructure of samples produced with set II parameters. Microstructure of set II samples also contained metallurgical defects such as pores and lack of fusion. Sporadically, coarse pores were observed. Size of coarse pores is approximately 100 µm in diameter. Comparing set I and set II samples, pores are more frequent in set I samples. Under higher magnification, difference in microstructure between center and edge of melt pool
At the edges microstructure is coarser, while moving away from edges, microstructure changes and becomes

finer (Figure 4b and d, Figure 5b and d) On XY plane microstructure is cell like, while on XZ/YZ plane microstructure is dendrite like (Figure 4f and h, Figure 5f and h).

4. DISCUSSION

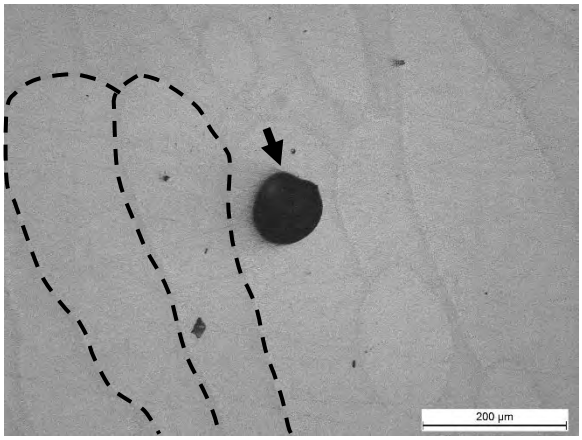
Visual examination revealed clear differences in surface quality between samples produced with parameter Set I and Set II. Samples produced using Set I parameters exhibited rougher surface, with presence of coarse spherical particles as well as fine powder particles. Additionally, delamination of layers was observed on samples produced using set I parameters suggesting inadequate interlayer bonding and lack-of-fusion defects. Formation of coarse spherical particles and delamination of layers can be attributed to balling effect. High reflectivity and thermal conductivity of Al can lead to formation of molten droplets on the powder bed surface. After solidification, these droplets form spherical particles which are known as balling phenomena [9]. The balling effect leads to uneven scan tracks and weak bonding between layers leading to porosity and delamination. Lack of balling defects and delamination of layers in set II parameter samples is attributed to higher energy density which provides enough molten material to reduce the occurrence of balling effect [9].

During LPBF process, high energy laser selectively melts metal powder which solidifies instantly as laser moves. This rapid solidification creates unique microstructure characterized by Al matrix and cellular network of Al-Si eutecticum. During solidification, high cooling rates lead to supersaturation of melt pool with silicon. As a result, microstructure exhibits fine Al-Si eutecticum. On Figure 4 and 5, two types of microstructures were observed: coarse and fine cell like and dendrite like. Coarse dendrite like structure grains, during cooling, nucleate and grow at the interface between previously solidified layer and melted phase of new layer. Subsequently, dendrites grow in the direction of heat flux [5, 10].

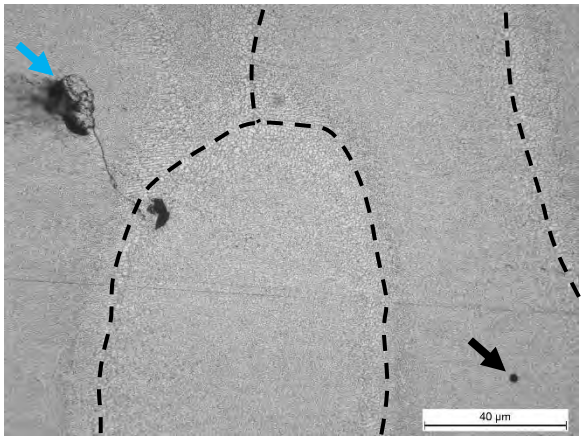


a) XY plane, magnification 100
b) XY plane, magnification 500
c) XZ/YZ plane, magnification 100
d) XZ/YZ plane, magnification 500
e) XY plane, magnification 100
f) XY plane, magnification 500
g) XZ/YZ plane, magnification 100
h) XZ/YZ plane, magnification 500

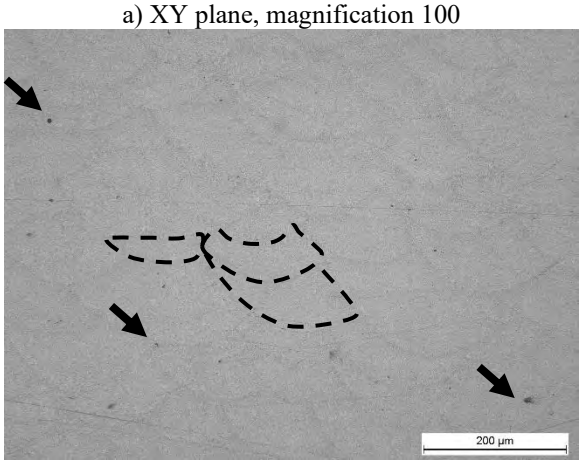
Figure 4. Microstructure of samples produced with Set I parameters; a-d) as polished, e-h) etched.



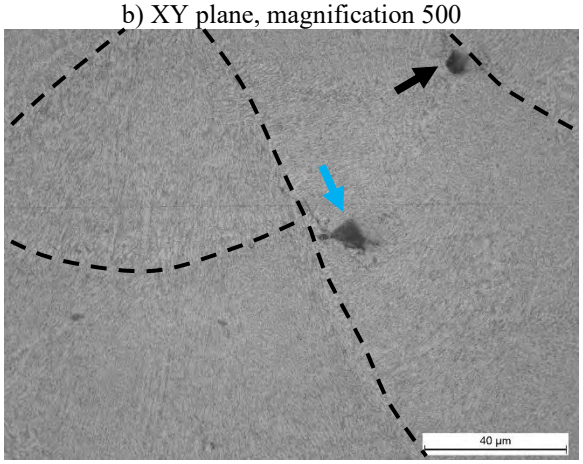
a) XY plane, magnification 100



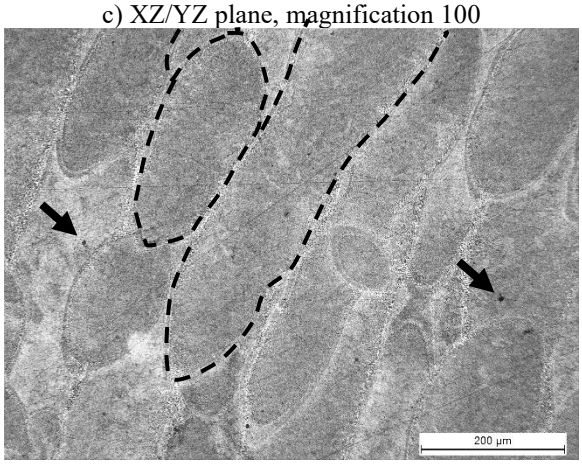
b) XY plane, magnification 500



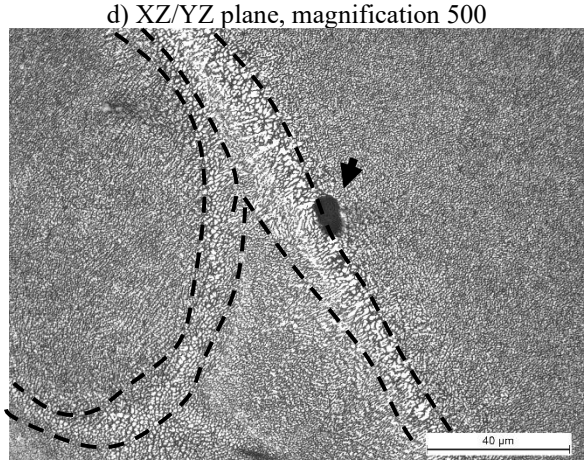
c) XZ/YZ plane, magnification 100



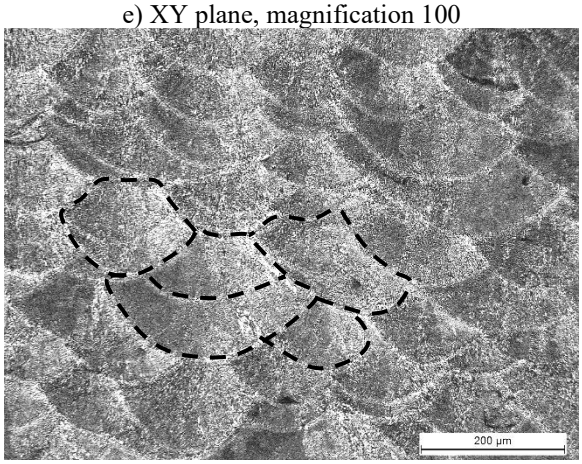
d) XZ/YZ plane, magnification 500



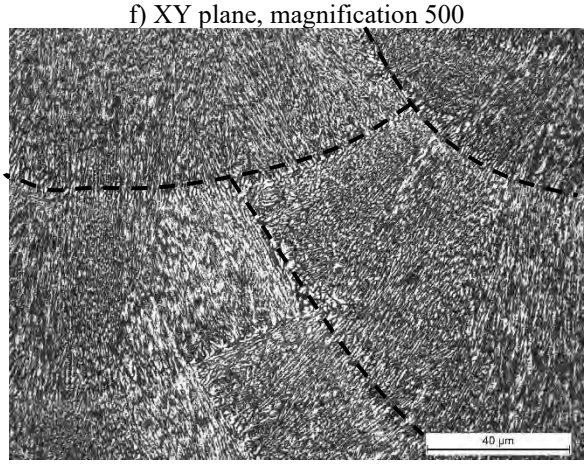
e) XY plane, magnification 100



f) XY plane, magnification 500



g) XZ/YZ plane, magnification 100



h) XZ/YZ plane, magnification 500

Figure 5. Microstructure of samples produced with Set II parameters; a-d) as polished, e-h) etched.

Chemical analysis performed using OES showed that both sets of samples had chemical compositions within the limits specified for EN AC-Al10SiMg alloy. No significant differences were observed between Set I and Set II samples.

Since samples from both sets of parameters were manufactured using powder from the same batch, the absence of compositional deviations confirms that differences in microstructure and defects are attributed to process parameters rather than metal powder variability.

Hardness values measured on XY and XZ/YZ planes showed no significant differences between the two sets of parameters. Set I samples exhibited mean hardness values of 130 HB (XY) and 127 HB (XZ/YZ), while Set II samples showed 130 HB (XY) and 119 HB (XZ/YZ). The results indicate that the difference in LPBF process parameters did not significantly affect bulk hardness. This is expected for AlSi10Mg alloy, where hardness is primarily governed by the fine cellular Al–Si eutectic formed under rapid LPBF cooling conditions. Since both sets of parameters resulted in fine eutectic structure, hardness remained unchanged [10].

5. CONCLUSION

This study investigates the influence of Laser Powder Bed Fusion (LPBF) process parameters on the quality of AlSi10Mg alloy samples. Two sets of samples were produced using different combinations of laser power, scan speed, hatch spacing resulting in two energy densities, 39.3 J/mm^3 for set I, and 50.3 J/mm^3 for set II parameters.

Samples manufactured with lower energy density (Set I) exhibited rougher surfaces, unmelted powder particles, and delamination between layers, indicating insufficient interlayer bonding and lack-of-fusion defects. Samples produced with higher energy density (Set II) showed smoother surfaces and no delamination. Microstructural analysis revealed typical LPBF features in both sets, including fine cellular Al–Si eutectic and coarse dendrite like grains at melt pool boundaries, with pores and lack-of-fusion defects present in all samples but more frequently in Set I samples.

Hardness measurements showed no significant differences between parameter sets or measurement planes. Overall, the results demonstrate that LPBF process parameters strongly affect surface morphology and defect formation, while having negligible influence on bulk hardness. Optimizing energy density is therefore essential for achieving uniform microstructure, adequate layer bonding, and improved part quality in AlSi10Mg components.

Acknowledgement

This work was supported by the Ministry of Science, Technological Development and Innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

Reference

[1] DONG Z., XU M., GUO H., FEI X., LIU Y., GONG B., JU G.. Microstructural evolution and characterization of AlSi10Mg alloy manufactured by selective laser melting, Jouranl of material research and technology,. 17 (2022) 2343-2354.

[2] ABOULKHAIR N. T., SIMONELLI M., PARRY L., ASHCROFT I., CHRISTOPHER T., HAGUE R.. 3D printing of Aluminium alloys: Additive Manufacturing of Aluminium alloys using selective laser melting, Progress in Materials Science, 106 (2019) 100578.

[3] LI W., LI S., LIU J., ZHANG A., ZHOU Y., WEI Q., YAN C., SHI Y.. Effect of heat treatment on AlSi10Mg alloy fabricated by selective laser melting: Microstructure evolution, mechanical properties and fracture mechanism, Materials Science and Engineering: A, 663 (2016) 116-125.

[4] MERTENS A., DELAHAYE J., DEDRY O., VERTRUYEN B., TCHUINDJANG J. T., HABRAKEN A. M.. Microstructure and Properties of SLM AlSi10Mg: Understanding the Influence of the Local Thermal History, Procedia Manufacturing, 47 (2020) 1089-1095.

[5] GHIO E., CERRI E.. Additive Manufacturing of AlSi10Mg and Ti6Al4V Lightweight Alloys via Laser Powder Bed Fusion: A Review of Heat Treatments Effects, Materials, 15(6) (2022) 2047.

[6] TANG M.. Inclusions, Porosity, and Fatigue of AlSi10Mg Parts Produced by Selective Laser Melting, PhD Thesis, Pittsburgh, 2017.

[7] GEBHARDT U., GUSTMANN T., GIEBELER L., HIRSCH F., HUFENBACH H., KÄSTNER M.. Additively manufactured AlSi10Mg lattices – Potential and limits of modelling as-designed structures, Materials & Design, 220 (2022) 110796.

[8] MĚSIČEK J., ČEGAN T., MA Q.-P., HALAMA R., SKOTNICOVÁ K., J. HAJNYŠ, JUŘICA J., KRPEC P., M. PAGÁČ,. Effect of artificial aging on the strength, hardness, and residual stress of SLM AlSi10Mg parts prepared from the recycled powder, Materials Science and Engineering: A, 855 (2022) 143900.

[9] BALLA S. K., MALLAIAH M., NAGAMUTHU S., DAS R. Overview of Laser Powder Bed Fusion and Laser Directed Energy Deposition of Aluminium Alloys: Process, Defects, Microstructure, and Mechanical Properties, Metals and Materials International, 32 (2026) 2584-2623.

[10] GADLEGAONKAR N., BANSOD P. J., LAKSHMIKANTHAN A., BHOLE K.. A Review on Additively Manufactured AISi10Mg Alloy: Mechanical, Tribological, and Microstructure Properties, Journal of Mines, Metals and Fuels, 73(1) (2025) 87-101.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EFFECT OF SIO_2 NANOPARTICLES ON THE MECHANICAL PROPERTIES OF POLYURETHANE THERMAL INSULATION LINERS

SARA PEPIĆ
Military Technical Institute, Belgrade, spepic00@gmail.com

JELENA GRŽETIĆ
Military Technical Institute, Belgrade, jelenarusmirovictmf@gmail.com

OLGA MARAŠ
Military Technical Institute, Belgrade, olgamaras98@gmail.com

VESELIN ŽIVANOVIĆ
Military Technical Institute, Belgrade, veselin.zivanovic.best@gmail.com

TEODORA STANČIĆ
Military Technical Institute, Belgrade, teodorastancic@gmail.com

Abstract: *This study investigates the development and characterization of thermal insulation composite liners (L) based on hydroxyl-terminated polybutadiene reinforced with carbon fibers of different lengths and modified with silicon dioxide nanoparticles (nano-SiO_2). The aim of this study was to examine the influence of nano-SiO_2 contents on the physicochemical and mechanical properties of the liners. The prepared composite liners were characterized by Fourier transform infrared (FTIR) spectroscopy, physicochemical analyses, and uniaxial tensile testing. The obtained results showed that the addition of nano-SiO_2 affects the dynamic viscosity of uncured composite liners, density and mechanical properties of the cured composite liners, indicating the importance of proper nano-filler dispersion and compatibility with the polymer matrix. This study provides a foundation for further optimization of liner formulations intended for rocket motor thermal protection systems. The results suggest that improved dispersion of nano-SiO_2 could further enhance the overall properties of the composites.*

Keywords: *composite liner, nano-SiO_2, dispersion, physical and mechanical properties*

1. INTRODUCTION

Thermal protection and structural integrity of solid rocket motors are highly dependent on the materials located between the propellant grain and the motor case. In composite solid rocket motors the propellant is bonded to the motor case through an elastomeric interfacial layer known as a liner. Positioned between the propellant and the motor case wall, the liner provides adhesion, transfers mechanical stresses and contributes to the overall integrity of the motor system [1].

Reliable bonding between the liner, propellant and motor case is essential for the proper operation of a solid rocket motor [2]. The liner acts as a structural interface responsible for stress transfer to the load-bearing components and prevents propellant grain movement. Loss of adhesion may lead to void formation, pressure fluctuations, unstable combustion, motor failure or explosion [3].

Therefore, the liner formulation must be chemically compatible with the propellant constituents while also being capable of establishing a stable chemical bond with them. To ensure such compatibility, the liner binder is typically based on the same polymeric system used as the binder in the propellant formulation [1]. In this study hydroxyl-terminated polybutadiene (HTPB) was used as the polymer matrix. HTPB is one of the most widely used binders in composite solid rocket propellants due to its favorable mechanical, thermal and chemical properties. It is characterized by a low glass transition temperature, pronounced hydrophobicity and a high concentration of terminal hydroxyl groups, which enable effective crosslinking with isocyanate curing agents forming a cross-linked polyurethane network.

Fibrous reinforcements, such as short Kevlar or carbon fibers, and flame retardants or intumescent additives, such as melamine or ammonium polyphosphate, improve the thermal properties of HTPB-based composite liners by forming a thermally protective char layer.

Moreover, liner composite formulations filled with micron-sized particles generally exhibit limited mechanical and functional properties. The development of nanocomposites, based on ultrafine fillers in the nanometer scale, enables significant improvements in mechanical, thermal and dynamic-mechanical performance compared to microfilled systems. The properties of these materials depend on the degree of cross-linking of the polymer matrix, the functional group

Corresponding author: S. Pepić
E-mail: spepic00@gmail.com
DOI: 10.5937/OTEH26132P

ratio (NCO/OH) and the type and content of fillers. Increased cross-linking density leads to higher stiffness and elastic modulus, while reducing elongation at break and swelling capacity. Therefore, functional fillers are used to optimize performance by simultaneously enhancing mechanical and barrier properties.

In this context, silicon dioxide nanoparticles (nano-SiO_2) is recognized as an effective filler due to its high specific surface area and excellent thermal stability. It contributes to improved mechanical properties and reduced migration of low-molecular-weight components [4]. However, its pronounced hydrophilicity may lead to particle agglomeration, making controlled dispersion within the polymer matrix essential [5].

The aim of this study was to investigate the effect of different nano-SiO_2 contents on the physicochemical and mechanical properties of HTPB-based thermal insulation liners containing carbon fiber reinforcement and flame-retardant additives for solid rocket motor applications. A four liner formulations containing various amounts of nano-SiO_2 were prepared and characterized. The influence of nano-SiO_2 concentration on thermal insulation liner performance was evaluated in order to identify the optimal filler content and to improve the overall material properties.

2. MATERIALS AND METHODS

2.1. Materials

For the preparation of thermally insulating liner formulations, several components were used. Hydroxy-terminated polybutadiene (HTPB) (average molecular weight, Mn: 2900 g·mol^{-1} and hydroxyl functionality: 2.4, glass transition temperature, T_g: -80 ºC) supplied by Evonik, Germany, was used as the polymer binder. Carbon fibers of different geometries were used as reinforcing fillers, including milled carbon fibers (0.2 mm) and chopped carbon fibers (3 mm), purchased from R&G Faserverbundwerkstoffe GmbH, Waldenbuch, Germany (Figure 1) and carbon nanofibers (CNFs) (purity ≥ 99.9%, 1-15 µm). Additional carbon fillers were used to improve the thermal stability and mechanical performance of the liner systems.

Ammonium polyphosphate (APP) and melamine (purity ≥ 99.8%) both procured by Thermo Scientific Chemicals, Germany, were incorporated as additives to enhance the thermal resistance and flame-retardant properties of the formulations.

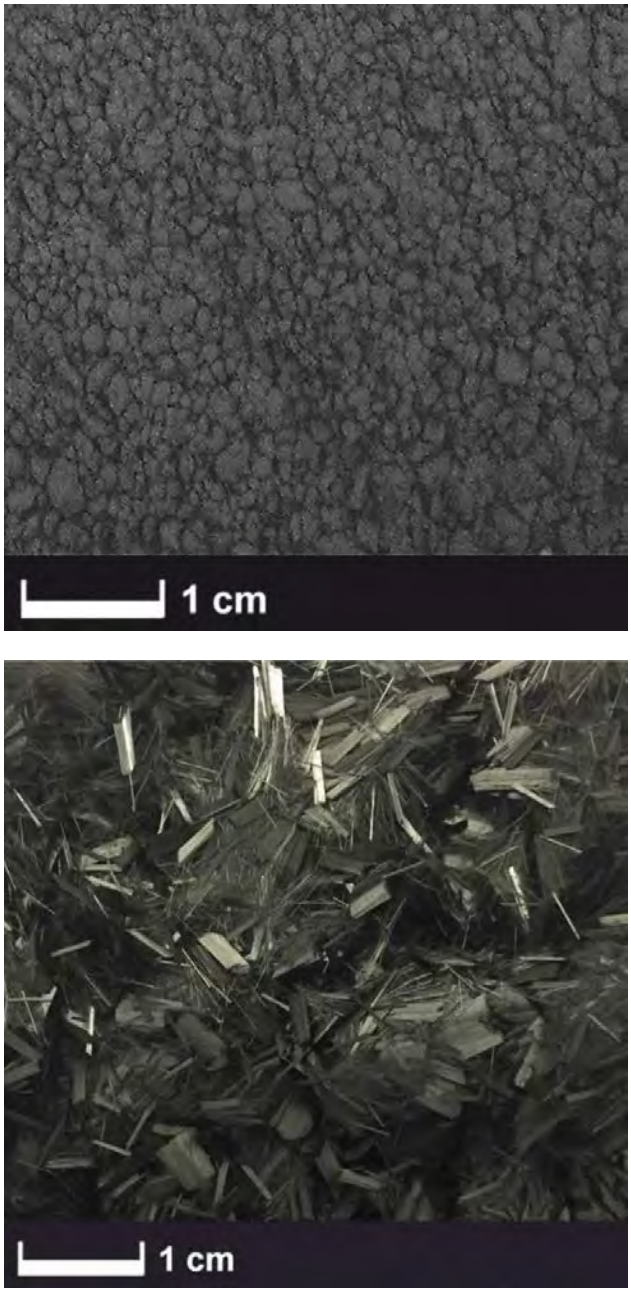


Figure 1. a) Milled carbon fiber 0.2 mm b) Chopped carbon fiber strands 3 mm (retaken from the R&G Faserverbundwerkstoffe GmbH)

Castor oil (*Ricinus communis seed oil,* functionality: 3, average molecular weight, Mn: 933 g·mol^{-1}) was used as a reactive plasticizer. The NCO/OH ratio in the formulations was adjusted to 0.33 to control the crosslinking density. The antioxidant AO 2246 (2,2'-bis(4-methyl-6-tertbutyl) phenol) supplied by Sigma Aldrich, Germany, was used to prevent oxidative degradation of the polymer matrix. Isophorone diisocyanate (IPDI) supplied by Evonik, Germany, while dibutyltin dilaurate (DBTDL) was used as a catalyst for the curing reaction. Nano-SiO_2 (purity ≥ 99.5%, particle size in the range of 20-30 mm), purchased by Speed UP (US Research), was used as a nano-filler to improve mechanical reinforcement and thermal stability of the liner systems.

2.2. Homogenization of the liners

Mixing and homogenization of the components (Table 1.) were carried out in a laboratory DRAIS mixer (1.5 L capacity) at room temperature and a stirring speed of 40 rpm. In the first stage, HTPB, RU and DBTDL were added into the mixing vessel. Subsequently, AO2246, APP, melamine, chopped and milled carbon fibers and nano-SiO_2 were gradually added in small portions under continuous mixing. Finally, the IPDI curing agent was added into the mixture and homogenization was continued at room temperature for an additional 15 minutes to ensure a uniform dispersion of all components. The liquid mixture was then cast into a PTFE mold (200 mm × 200 mm × 3 mm) and cured in an oven for four days at 60 ºC.

Table 1. Composition of liners

Component	Mass content (wt. %) of liner formulations			
	L1	**L2**	**L3**	**L4**
HTPB	36.57	36.77	36.89	36.97
Castor oil	5.85	5.88	5.90	5.91
AO 2246	0.18	0.18	0.18	0.18
Melamine	25.00	25.00	25.00	25.00
APP	25.00	25.00	25.00	25.00
CNFs	0.50	0.50	0.50	0.50
Carbon fibers; 0.2 mm	2.40	2.40	2.40	2.40
Carbon fibers; 3.0 mm	0.50	0.50	0.50	0.50
Nano-SiO_2	**0.50**	**0.25**	**0.10**	**-/-**
DBTDL	0.06	0.06	0.06	0.06
IPDI	3.44	3.45	3.47	3.47

2.3. Characterization of Materials

The remaining amount of the uncured liner was used for the determination of dynamic viscosity using Brookfield HBT viscometer at 25 ºC. Measurements were performed during the first 75 minutes after completion of the mixing procedure and incorporation of all components.

The Fourier transform infrared spectroscopy (FT-IR) spectrum of the cured liners was recorded in absorbance mode using a Nicolet™ iS™ 10 FT-IR spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), within a range of 400-4000 cm^{-1} at a resolution of 4 cm^{-1}.

The density of the cured liner formulations was determined using a KERN YDB-03 analytical balance (Version 1.4; 2021-003) at 23 ºC.

For the evaluation of mechanical properties, test specimens were prepared with dimensions of 115 x 25 x (3 ±1) mm and a gauge length of 46.5 mm. Mechanical testing was performed on an Instron 1122 Universal Testing Instrument (Instron, Nortwood, MA, USA) at 20 ºC with a loading rate of 50 $mm \cdot min^{-1}$.

The following mechanical properties were determined:

- tensile strength (σ_m),
- elongation at maximum force (ε_m), and
- elongation at break (ε_p)

3. RESULTS AND DISCUSSION

3.1. FTIR Analysis

FTIR analysis was used to evaluate the chemical structure of the prepared liner formulations and to confirm the formation of the polyurethane network during the curing process. The FTIR spectra of samples L1 – L4 are shown in Figure 2. The absorption bands at 3463 and 3516 cm^{-1} are assigned to N-H stretching vibrations characteristic of polyurethane structures, while the weak band at 3322 cm^{-1} indicates the presence of residual hydroxyl groups that did not fully react during network formation. The absorption bands at 2916 and 2845 cm^{-1} correspond to aliphatic (-CH_2-) stretching vibrations of the HTPB backbone. The characteristic band at 1706 cm^{-1} corresponds to the stretching vibration of the urethane carbonyl group (C=O), confirming the formation of the polyurethane structure. The band at 1640 cm^{-1} is assigned to C=C stretching vibrations of residual unsaturated bonds within the polybutadiene chains.

Bands at 1516 and 1435 cm^{-1} are attributed to N-H bending and C-H bending vibrations, respectively, while the peak at 1238 cm^{-1} corresponds to C-N stretching vibrations within the urethane structure. In the lower wavenumber region, absorption bands at 964 cm^{-1} and 910 cm^{-1} confirm the presence of Si-O-Si and Si-O vibrations, confirming the presence of nano-SiO_2 particles within the polyurethane matrix.

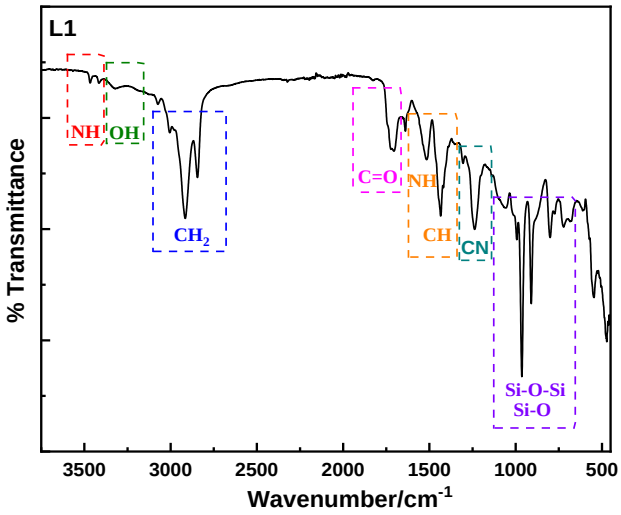


Figure 2. FTIR analysis of the liner L1

3.2. Viscosity measurements of uncured composite liners

After the homogenization process, the liner must exhibit an appropriate viscosity to ensure proper casting and application onto the inner wall of the rocket motor chamber. The time during which the material retains suitable rheological properties for processing is defined as the gelation time – *pot life*. This parameter depends on the resin-curing agent system, as well as on the type and content of fillers and other additives.

The viscosity of uncured liner formulations changes over

time and represents a key parameter for *pot life* determination. Therefore, dynamic viscosity was measured at 15-minute intervals after homogenization, over a total period of 75 minutes, which was considered sufficient for the planned application procedure.

The results are presented in Table 2, while the corresponding graphical representation is shown in Figure 3.

Table 2. Dynamic viscosity of uncured liner formulations.

Time (min)	Dynamic viscosity (Pa·s)			
	L1	L2	L3	L4
15	592	624	736	800
30	736	816	752	1072
45	/	/	800	1312
60	912	1088	896	/
75	/	1184	992	/

Note: L1-L4 represent different liner formulations with different contents of nano-SiO_2

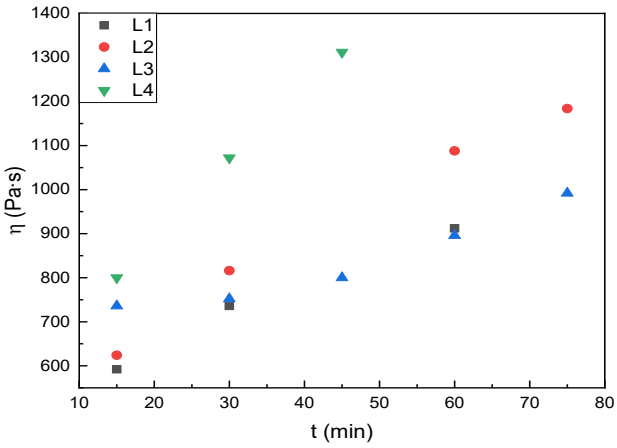


Figure 3. Comparison of the initial viscosity of experimental liner formulations

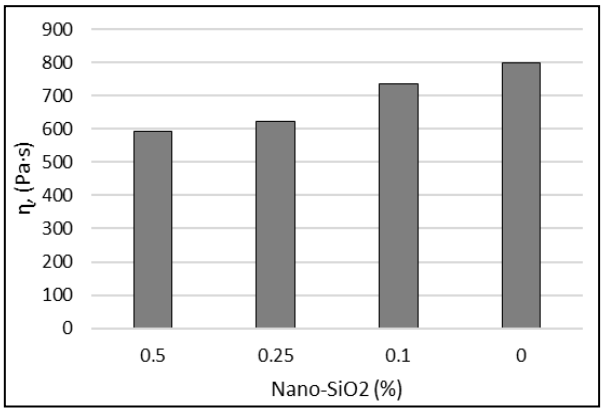


Figure 4. Dynamic viscosity 15 minutes after IPDI addition

Based in the time-dependent viscosity profiles, an increase in viscosity over time is observed, confirming the progress of crosslinking reactions. The results show that sample L4, without nano-SiO_2, exhibits the fastest and most pronounced viscosity increase (Figure 4). With increasing nano-SiO_2 content, the initial viscosity is partially reduced. However, sample L1, containing the highest nano-SiO_2 content (0.5 wt.%), shows the lowest viscosity after 15 minutes from the addition of the curing agent.

The decrease in viscosity with increasing nano-SiO_2 content is attributed to nano-particle dispersion within the HTPB/IPDI matrix. Such insertion of nano-SiO_2 between HTPB polymer chains and IPDI cross-linker partially interrupts the crosslinking reaction. The gradual viscosity decrease at higher nano-SiO_2 contents extends the gelation time and improves control of the liner application process.

3.3. Density determination

A key requirement for thermal insulation liners is low density (1.05-1.10 g·cm^{-3}) to reduce the overall mass of the propellant grain. The addition of fillers requires a balance between improved thermal stability and ablation resistance and the resulting increase in density. High-density fillers such as silicon dioxide contribute to the formation of a more stable char layer. However, their content must be optimized to maintain low liner density. The average density values of the prepared formulations are presented in Figure 5. A slight increase in density with increasing nano-SiO_2 content was also confirmed.

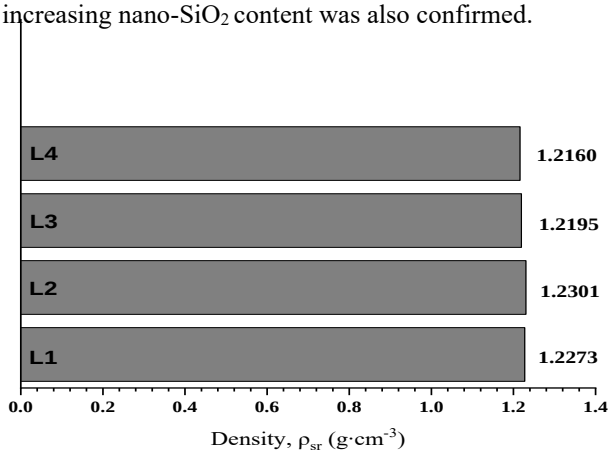


Figure 5. Density values of liner formulations L1-L4.

Note: The reported values are the average of four measurements.

3.4. Mechanical characterization

Mechanical testing was performed to evaluate the effect of nano-SiO_2 content on the tensile properties of the developed liner formulations (Figure 6). The obtained values of tensile strength, elongation at maximum force, elongation at break and Young's modulus (Table 3) indicate that the incorporation of different amounts of nano-SiO_2 significantly affects the mechanical behavior of the liners.

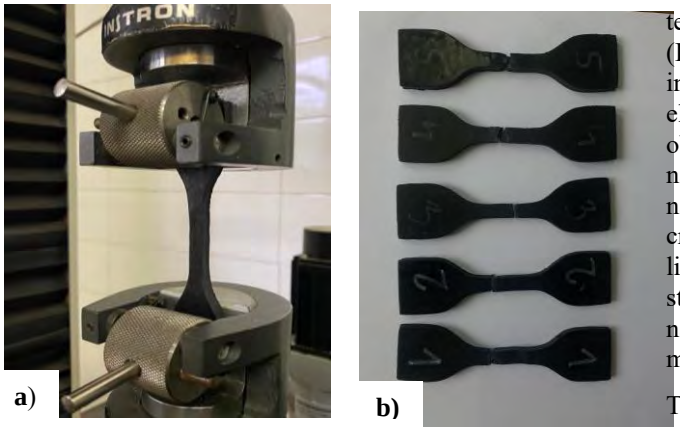


Figure 6. Appearance of L2 specimens before (a) and after (b) tensile strength testing

The polymer structure and mechanical performance of polyurethane-based liners are directly related to the degree of cross-linking achieved during the curing process. Highly cross-linked networks generally exhibit increased stiffness and tensile strength, while lower cross-linking densities promote greater flexibility and elongation. The tested formulations demonstrated typical elastomeric behavior characterized by large reversible deformations under tensile loading.

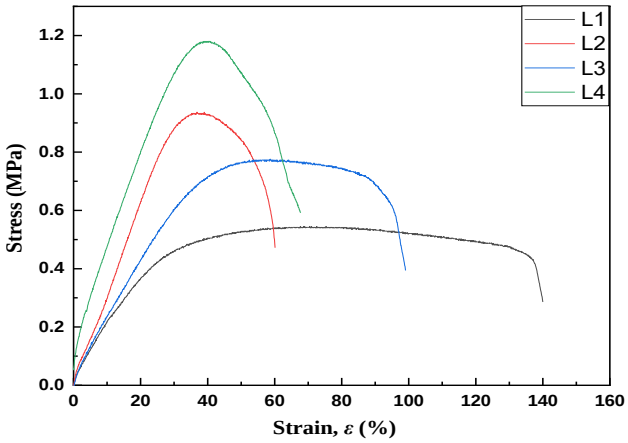


Figure 7. Stress-strain curves of the tested materials L1-L4

Table 3. Tensile strength, deformation at maximum force and at break and Young's modulus of the tested materials at T=20 ºC.

	σ_m, [daN/cm^2]	ε_m, [%]	ε_p, [%]	E [MPa]
L1	5.66 ± 1.30	51.97 ± 14.77	116.23 ± 23.68	1.08 ± 0.02
L2	9.34 ± 1.14	39.59 ± 8.25	66.85 ± 15.98	3.98 ± 0.02
L3	8.30 ± 1.34	52.54 ± 11.98	90.50 ± 29.54	3.98 ± 0.01
L4	10.75 ± 1.51	49.75 ± 10.78	79.22 ± 27.72	3.98 ± 0.01

Note: The reported values are the average of five measurements.

Among the investigated samples, L1 containing the highest nano-SiO_2 content (0.5 wt.%) exhibited the lowest tensile strength and the highest elongation at break (Figure 7). As the nano-SiO_2 content decreased, a gradual increase in tensile strength accompanied by a decrease in elongation at maximum force and elongation at break was observed. This behavior suggests that higher amounts of nano-SiO_2 contribute to the more pronounced insertion of nano-SiO_2 between HTPB polymer chains and IPDI cross-linker which significantly interrupts the cross-linking reaction. Moreover, the reduction in tensile strength may be attributed to the tendency of silica nanoparticles to form agglomerates within the HTPB matrix.

The observed decrease in tensile strength with increasing nano-SiO_2 content is consistent with literature data and can be primarily associated with nanoparticle agglomeration [7]. At lower concentrations, nano-SiO_2 particles are more uniformly distributed within the polymer matrix, whereas increasing particle content promotes particle-particle interactions and the formations of larger agglomerates. These agglomerates act as stress concentration sites and facilitate crack initiation during tensile loading, leading to a reduction in tensile strength Furthermore, the optimal values of tensile strength and elongation at break and maximal force were obtained for sample L3 indicating an optimal homogenization condition and nano-SiO_2 distribution.

Overall, the obtained results indicate that the mechanical performances of the developed liners are strongly dependent on the dispersion state of nano-SiO_2 and the resulting structure of the polyurethane network. Surface modification of nano-SiO_2 with reactive organosilane coupling agents, such as amino-, epoxy-, or isocyanate-functionalized silanes, represents a promising strategy for improving nanoparticle dispersion and overcoming compatibility limitations.

4. CONCLUSION

The influence of nano-SiO_2 and carbon nanofibers on the properties of HTPB-based thermal protection liners was investigated. FTIR analysis confirmed the formation of polyurethane bonds and successful curing of all developed formulations. Rheological measurements showed that the incorporation of nano-SiO_2 affected the viscosity evolution and *pot life* of the liner mixtures, indicating its influence on the curing process and network formation.

Mechanical testing revealed that increasing the nano-SiO_2 contend led to a decrease in tensile strength and an increase in elongation at break. The highest tensile strength was obtained for the formulation without nano-SiO_2 content exhibited the greatest elongation. These results were attributed to the tendency of nano-SiO_2 particles to agglomerate within the polymer matrix, affecting the degree of crosslinking and the homogeneity of the polyurethane network.

The obtained results demonstrate that the content and dispersion of nano-SiO_2 play a crucial role in determining the processing characteristics and mechanical performance of HTPB-based liners. Further optimization

of nanoparticle dispersion and surface modification of nano-SiO_2 may provide additional improvements in liner properties and contribute to the development of advanced thermal protections materials for rocket motor applications.

Acknowledgments

This work was supported by the Ministry of Science, Technological development and Innovations (Serbia), (Contract No. 451-03-34/2026-03/200325).

References

[1] RODIĆ, V: Case bonded system for composite solid propellants, Naučno-tehnički pregled, 57, 2007.

[2] SANDEN, R. WINGBORG, N.: Case bonded liner, Journal of Applied Polymer Science, (1989), Vol. 37, pp. 167-171

[3] SKIDMORE, P.H.: Case bonded system for cast composite propellants, US Pat. 3,813,308, May 28, 1974.

[4] PETROVIĆ, Z. S: Effect of silica nanoparticles on morphology of segmented polyurethanes. Polymer, (2004), 45, 4285-4295.

[5] WANG, X., WANG, L., Su, Q., ZHENG, J: Use of unmodified SiO_2 as nanofiller to improve mechanical properties of polymer-based nanocomposites. Composites Science and Technology, 2013.

[6] BHUVANESWARI, C. M., SURESHKUMAR, M. S., KAKADE, S. D., GUPTA, M: Ethylene propylene diene rubber as a futuristic elastomer for insulation of solid rocket motors. Defence Science Journal 56, 2006.

[7] POREBA, R., SPIRKOVA, M., & HRDLICKA, Z: Mechanical and thermomechanical properties of polycarbonate-based polyurethane-silica nanocomposites. Processing and Application of Ceramics, 2011.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

NUMERICAL ANALYSIS OF THE HOT FORGING PROCESS FOR A STEEL DEFENSE INDUSTRY COMPONENT

NIKOLA KASTRATOVIĆ
Zastava Kovačnica A.D. Kragujevac, Kosovska 4b, 34000 Kragujevac, Serbia; nikolakastratovic93@gmail.com

VESNA MANDIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; mandic@kg.ac.rs

MARKO DELIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; marko.delic@kg.ac.rs

DUŠAN ARSIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; dusan.arsic@fink.rs

NIKOLA KOMATINA
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; nkomatina@kg.ac.rs

Abstract: *The paper deals with the numerical simulation of the hot forging process on a hammer. Today, it is hard to think of reliability in automotive, mechanical, or defense engineering without using forged parts. Their toughness and field safety are just too critical. But to get a high-quality component, you have to know exactly how the metal behaves and flows inside the die during heavy deformation. This is why numerical process simulations have become our go-to tool for updating or creating new forging technologies.*

Instead of burning budget on slow and expensive shop-floor trials, we built a finite volume model to map out deformation, shifting stresses, and forging forces while the steel is compressed under the hammer. We focused heavily on pinning down those weak spots in the tooling that can mess up the final geometry or trigger unexpected downtime.

The data from Simufact Forming basically gave us a clear window into the material's core during the hits. It let us judge whether the planned workflow actually works, showing us exactly how the die fills and how efficient the whole setup is. This CAE-driven method makes perfect sense for modern manufacturing because it cuts down product development time drastically.

Keywords: *hot forging, Simufact Forming, material flow, defense indsutry, numerical simulation*

1. INTRODUCTION

Hot forging is one of the most important technologies for metal forming. Modern technologies such as 3D metal printing have been developed, but hot forging plays a very important role in the manufacturing process of parts under high loads. Hot forging has an important role in military and automotive industries due to good mechanical properties of manufactured parts. The lower values of the yield stress are obtained when the workpiece is heated up. However, it is very difficult to lubricate because of the high temperatures. In hot forging process, the tools are loaded by high temperature, friction due to sliding of material on tool surface and in case of hammer forging, impact load.

Numerical simulations of processes using software based on the finite volumes method are successfully applied to eliminate material flow defects, determine and optimize process parameters, determine tool stresses to prevent tool breakage, and predict workpiece properties through stress-strain analysis. In addition, it is very easy to make changes to the tool design and monitor the impact of these changes on process parameters, thus achieving optimal tool geometry at all stages of the process.

Die life acts as an important factor on the productivity. In warm forging process, wear is the predominant factor affecting the die life since it sustains higher temperature and working pressure. Lee and Jou [1] reported that die life in warm forging is affected by wear in over 70%. This kind of failure affects not only the cost but also the tolerances of the forging. Developing methods which allow prediction of wear and which are suited to be used in the design stage is the most effective way in order to optimize the process.

In accordance with the methodological approach of Lee and Jou [1], the lifetime estimation of hot forging tools requires the application of Archard's wear law integrated within a finite volume numerical simulation. This

Corresponding author: N. Kastratović;
E-mail: nikolakastratovic93@gmail.com
DOI: 10.5937/OTEH26133K

approach enables a precise correlation between the local conditions at the interface (pressure, sliding speed, temperature) and the predicted wear in the most stressed zones of the engraving.

According to Hartley and Pillinger [2], the degradation of finite elements is one of the primary challenges in the simulation of the forging process, which is why the implementation of reliable re-meshing strategies is necessary to maintain the stability of the calculation. Accordingly, in this paper, a nonlinear refresh criterion based on an increase in effective plastic deformation of 0.4 was adopted.

The phenomena at the interface between the forging tool and the workpiece are extremely complex and involve the combined influence of high friction, heat generation and material flow. When defining the numerical model in this paper, the influence of imperfect lubrication was simulated using a friction factor of 0.65, which required additional optimization of the size of the final mesh size to 1.5 mm in order to avoid numerical anomalies.

For the numerical analysis of the hot forging process, the Simufact Forming software was used, which in recent research (Delić et al) [3] has been successfully applied for the prediction of tool wear and the analysis of deformation forces. The numerical model in this paper was additionally validated through the optimization of flow parameters under highly loaded friction conditions [3].

As pointed out by Hawryluk et al. [4], tribological conditions in the hot forging process are extremely dynamic and subject to changes due to lubricant degradation. In accordance with these findings, the friction factor μ=0.65 was applied in this paper, which covers realistic conditions of limited lubrication at the tool-workpiece interface.

Following the methodology applied by Davoudi et al. [5], numerical analysis of forging tools requires precise definition of local contact stresses. The application of fine discretization (mesh size 1.5 mm) in this work enabled the detection of critical zones of stress concentration on the engraving radii, which ensures high reliability of the prediction of tool behavior under the influence of large deformation forces.

Yin et al. [6], emphasize that incorrect positioning of the ingot (billet) in relation to the center of the die leads to eccentric loading and premature breakage of the central punch, which can be successfully corrected by optimizing the guides and the forging positioning process itself.

As Řiháček et al. [7] proved in their comparative analysis of numerical methods in the Simufact.forming software, the finite volume method (FVM) shows superior numerical stability compared to the classical finite element method (FEM) when modeling processes with large deformations and intensive material flow.

The network refresh strategy implemented in this paper, based on the plastic strain increment of 0.4, is fully aligned with the findings of Řiháček et al. [7], who point out that optimal accuracy and stability of calculations are achieved by a combination of a suitable numerical solver and controlled adaptation of the network model.

As pointed out by Chenot et al [8], numerical simulation is increasingly replacing expensive and time-consuming trial and error methods on real machines, allowing engineers to find optimal solutions faster. The integration of optimization modules with simulation software allows not only to reduce production costs, but also to improve the mechanical properties of the final part through controlled preform geometry.

2. SIMULATION CONDITIONS

A three-dimensional thermo-mechanical finite volumes model was developed using Simufact Forming software in order to analysis the upset forging process of an industrial component manufactured from 42CrMo4 steel.

Steel 42CrMo4 (according to the standard EN 10083) is a low alloy improvement steel, which is very appreciated in the forging industry because of its high strength, excellent toughness and good hardenability. The material contains a precise balance of chromium and molybdenum, resulting in superior resistance to material fatigue and impact loads and is the ideal choice for the manufacture of components that are highly stressed in the defense and automotive industries. 42CrMo4 presents a stable behavior at high temperatures during the hot forging process, which allows to obtain complex geometries without the risk of internal defects, and after a proper thermal treatment it presents superior mechanical properties necessary to work in the most difficult operating conditions.

The simulation model consisted of three main geometrical entities: the workpiece, the upper die and the lower die. The workpiece geometry was imported as a CAD model and discretized using an adaptive size of mesh.

The workpiece material was defined as $42CrMo_4$ steel using the material data available in the Simufact material database. The material behavior was considered temperature-dependent in order to accurately represent the hot forging conditions.

First case: The forging operation was simulated on a Hammer MPM-1000 machine. The initial workpiece temperature was set to 1200°C, while the initial die temperature was 250°C. The ambient temperature was defined as 20°C. Contact interactions between the workpiece and the dies were modelled using friction conditions 0.5.

Second case: The forging operation was simulated on a Hammer MPM-1000 machine. The initial workpiece temperature was set to 1100°C, while the initial die temperature was 250°C. The ambient temperature was defined as 20°C. Contact interactions between the workpiece and the dies were modelled using friction conditions 0.65.

The simulation was performed for the given thermo-mechanical conditions, considering the heat transfer between the workpiece and the tools, the temperature increase due to plastic deformation and the influence of temperature on the material flow behavior.

The aim of the numerical model was to analyze the material flow, distribution of the effective strain, stress state and development of the forging force during forming operation. The obtained results were used to evaluate the suitability of the selected technological solution and to indicate potential critical regions that could influence the quality of the final forging.

3. RESULTS AND DISCUSSION

The numerical simulation provided a detailed insight into the thermo-mechanical behavior of the workpiece during the hot die forging process. The obtained results enabled the evaluation of material flow, effective strain distribution, stress state, temperature field and forging force development throughout the forming operation.

Two numerical simulations of the hot die forging process were performed in order to investigate the influence of friction conditions on the material behavior during forming.

The first simulation was carried out using a shear friction factor of 0.50, while the second simulation considered an increased friction factor of 0.65. In addition, a finer mesh size was applied in the second case.

The obtained results showed that the increase in friction reduced the maximum effective strain from approximately 6.25 to 4.52, indicating a more restricted material flow. At the same time, local material compression increased, which was reflected in higher local density values and a greater number of remeshing operations during the simulation.

The maximum workpiece temperature remained within the expected range for both cases, although a slight reduction was observed for the higher friction condition.

The numerical results confirm that friction conditions significantly affect material flow behavior, deformation distribution and overall process stability during hot die forging.

The analysis confirmed that the material successfully filled the die cavity under the selected technological conditions. The distribution of effective plastic strain and stresses indicated the regions subjected to the highest deformation intensity, while the temperature field showed the thermal conditions during the forming process.

Special attention was paid to the influence of friction conditions on the material behavior. Comparative analysis of the investigated cases demonstrated that the change in friction coefficient affected the material flow pattern, deformation distribution and contact conditions between the workpiece and the dies.

In addition to the standard process parameters, the numerical model enabled the evaluation of potential critical regions by means of the Cockroft-Latham damage criterion and die wear prediction. These results provide useful information for assessing process stability and the quality of the final forging.

Increase in wear: When moving from μ =0.5 to μ =0.65, the maximum die wear (Max die wear) increased from 1.3E-11 to 8.54E-11. This is almost 6.5 times the higher value, confirming that even small changes in the tribological conditions at the tool-workpiece interface drastically shorten tool life.

Drastic jump in pressure: The Max pressure value jumps from 4691.77 MPa to an incredible 22829.8 MPa Such high pressure values indicate that with greater friction, the material moves much harder through the forging engraving, which creates an extreme load on the surface of the tool.

Deformation parameters: It is interesting that the Max effective plastic strain decreases from 6.25 to 4.52 at higher friction. This is the expected behavior because higher friction restricts the free flow of the material which directly confirms the theoretical proposition that friction dictates the deformation mechanism in hot forging.

Table 1. Comparation of results

Parameter	*μ=0.5*	*μ=0.65*
Max effective plastic strain	6.25	4.52
Max pressure	4691.77 MPa	22829.8 MPa
Max plastic strain rate	29235.5	194846
Max die wear	1.3E-11	8.54E-11
Max temperature	1286.11°C	1263.25°C

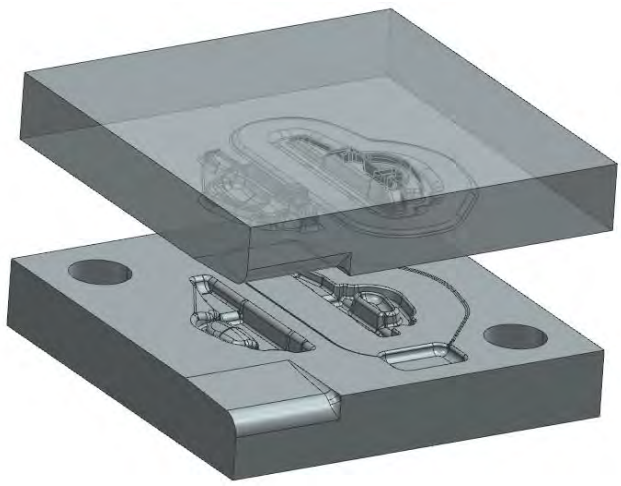


Figure 1. Upper and lower die for forging

Figure 1 shows a 3D model of the tool assembly, which includes the upper and lower dies imported in the Simufact Forming software environment. The geometry of the tool is imported from the CAD system with the application of precise discretization, which ensures a reliable mapping of contact surfaces and interactions with the workpiece during the forging process.

The forging process of the analyzed industrial component consists of operations: the preforming operation and the finishing operation, which are necessary to achieve the required geometric characteristics and precision of the

part. The initial step includes the operation of compressing (upsetting) the preform, which prepares the raw material for the deformation process, followed by the final stage, in which the material is completely pressed into the engraving of the tool under the influence of high pressures and temperatures. This sequential division into preparatory and final operations enables a more even material flow, reducing the risk of internal defects and optimizing the use of energy during hammer forging.

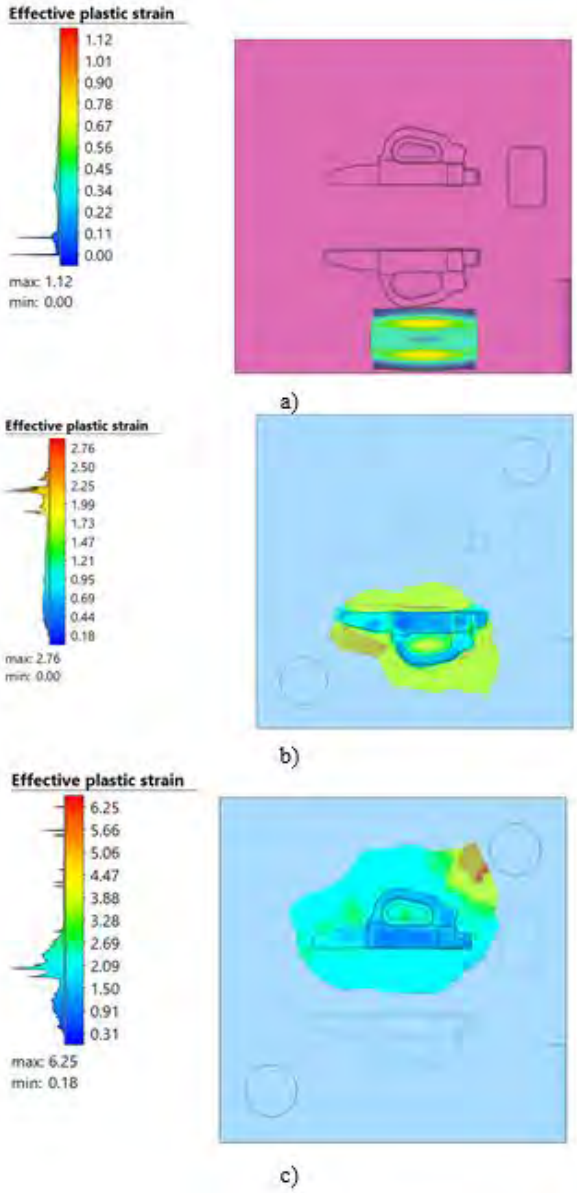


Figure 2. Value of Plastic Strain for first case: a) Upsetting; b) Pre-forging; c) Final forging

Figure 2 shows the distribution of the effective plastic deformation in the final phase of the forging hammer stroke (at 65 mm stroke) for the initial numerical model with a defined friction factor μ = 0.5. The contour display indicates a characteristic inhomogeneous deformation field, typical for hot forging processes of asymmetric parts in open molds. The maximum value of the effective plastic deformation reaches a value of 6.25 and is strictly localized on the periphery of the forging, in the zone of crown formation (upper right segment figure 3c). This high value is a direct consequence of the intense shearing of the material that leaves the central engraving under pressure and enters the narrow rim channel. This peripheral ring of highly deformed material creates the necessary hydrostatic resistance, which forces the remaining volume of 42CrMo4 steel to fill the deepest segments of the tool, such as the trigger guard zone.

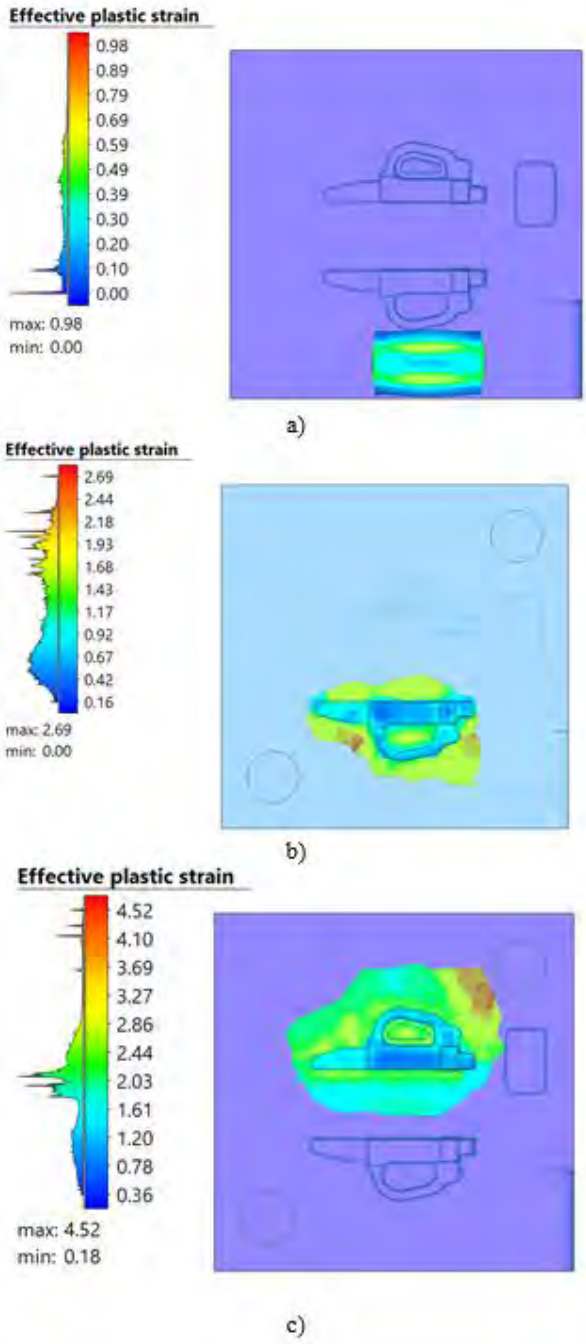


Figure 3. Value of Plastic Strain for second case: a) Upsetting; b) Pre-forging; c) Final forging

In contrast to the rim zone, the working part of the forging (central zone) shows a significantly lower and more homogeneous deformation profile. This uniform distribution in the core of the component indicates a stable material flow without the appearance of critical zones that could lead to internal defects, such as folds or cracks. However, although the deformation field in the central part is stable, this basic network model showed signs of numerical instability at the very end of the stroke

(Increment 2143). This served as the basis for further optimization of the remeshing conditions and the size of the final elements in subsequent iterations, especially in analyzes with more realistic, higher friction factors.

In order to achieve more realistic forging conditions, which include imperfect lubrication and high flow resistance in contact with the tool, an optimized numerical model with a friction factor of μ = 0.65 was created. Figure 4 shows the distribution of effective plastic deformation at the very end of the process (at 65 mm stroke). In order to preserve the stability of the calculation under such rigorous conditions, a fine spatial discretization was applied (mesh size of 1.5 mm) with a stricter nonlinear criterion of mesh refresh (remesh) at a plsatic strain increment of 0.4.

Unlike the base model, the maximum recorded effective plastic strain value is now 4.52. This drop in the maximum value is an indicator of the exceptional stability of the solver; namely, the finer mesh and stricter remesh criterion successfully eliminated local numerical anomalies and "lock-in" effects at sharp radii, enabling smooth mathematical tracking of material flow.

As expected, the zone of maximum deformation (red spectrum) is still localized in the peripheral ring of the forging part. The main progress of this model is, however, presented in the centre, working zone of forging. The deformation field in the region of the rifle body and trigger guard is significantly more homogeneous.

This stable and visually continuous material flow under high friction conditions proves the robustness of the applied meshing strategy. The obtained results confirm that the optimized model is free from non-physical peaks, thus providing highly reliable initial data for any further thermo-mechanical analysis of forging tools and the assessment of their service life.

The obtained simulation results confirm that numerical analysis represents an efficient tool for the investigation and optimization of industrial forging processes. The developed model made it possible to evaluate the influence of selected technological parameters without the need for costly experimental trials.

The comparison of the analyzed cases indicates that friction conditions have a significant influence on the material flow and deformation behavior during the forging process. Variations in friction affect the contact conditions between the dies and the workpiece, which consequently influence the stress distribution, effective strain and overall process performance.

The application of damage and die wear criteria contributes to a better understanding of potential critical zones and provides additional information for improving tool design and technological parameters.

The presented approach demonstrates that CAE methods can support the development of industrial forging technologies by reducing development time, improving process reliability and contributing to higher product quality.

For a more detailed analysis (Figure 4 and 5) of the state of deformation, mapping of the effective plastic deformation was carried out using measuring points placed at characteristic places inside the working part of the forging. This approach made it possible to precisely monitor the intensity of deformation in the zones of interest, which eliminates the influence of numerical extremes localized on the rim and gives a more realistic insight into the distribution of deformation within the piece itself.

Case 1 ($\mu=0.5$), Figure 4: Application of a higher friction coefficient resulted in a more homogeneous deformation profile in the same zones, with values ranging from min 0.38 to max 2.98. The reduction of the maximum deformation values in the working part of the forging confirms that the decreased friction limits the free flow of the material, which achieves a more stable filling of the engraving and reduces the risk of internal defects due to uneven flow.

Case 2 ($\mu=0.65$), Figure 5: In the central zones of the forging, recorded values of effective plastic deformation range from min 0.36 to max 3.37. This distribution indicates a more intensive penetration of the material into the engraving of the tool, but also potentially greater variations in the deformation field due to bigger frictional resistance.

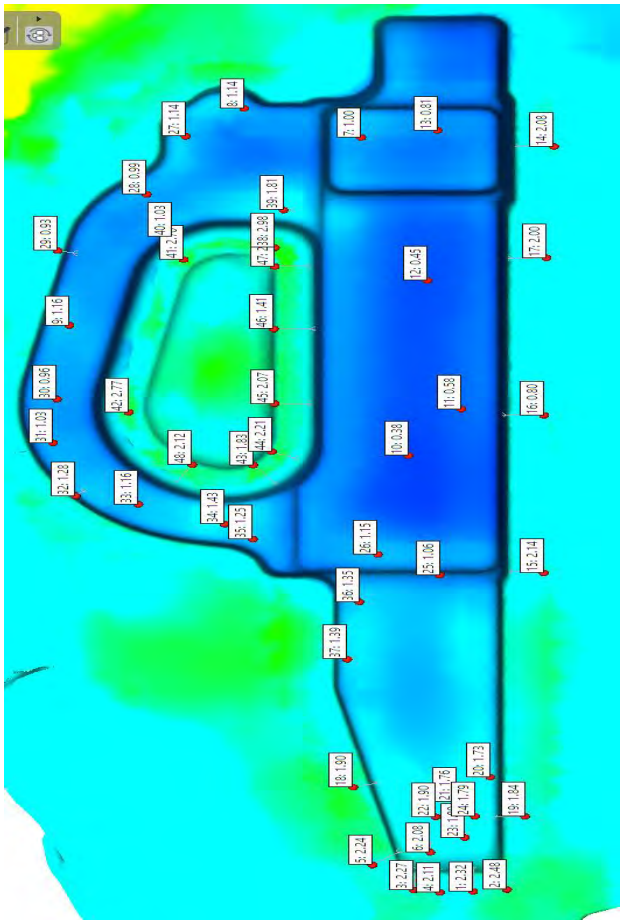


Figure 4. Plastic Strain point on forging part first case

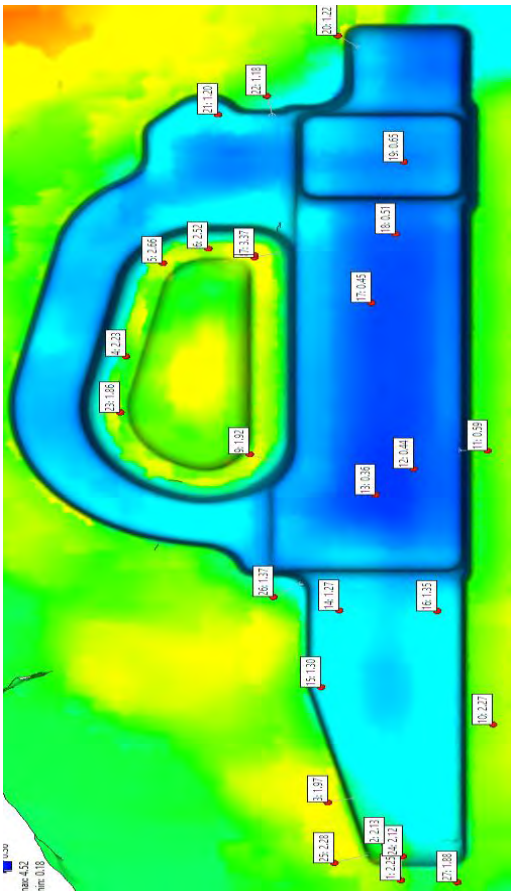


Figure 5. Plastic Strain point on forging part second case

5. CONCLUSION

In this paper, a numerical analysis of the forging process was carried out, where the finite volume method (FVM) was applied and showed exceptional stability and reliability in the simulation of large plastic deformations. The comparative analysis indicates that the results obtained by the FVM solver are fully comparable with the experimental data, thus confirming their applicability in industrial tool design. The key contributions of the research are reflected in the following: It has been confirmed that the precise definition of friction conditions, using the shear model with the factor μ =0.65, is the basis for a reliable simulation of wear. The application of the Simufact wear model enables the early detection of critical zones on the tool, which directly affects the increase of the service life of the dies. The analysis showed that the independent control of the movement of the punch significantly reduces the risk of material defects, such as "cold joints" (cold shut), which significantly improves the quality of the final product. Based on the above, it can be concluded that the integration of modern numerical simulations and experimental verification provides a comprehensive tool for forging process optimization. This approach enables the reduction of development costs, faster production and guaranteed high-quality production of parts with complex geometry.

The results of this study show that the conditions of friction have a major influence on the hot forging process. A higher friction factor results in a more restricted material flow, which increases the contact pressures at the interface between the tool and the workpiece, and in turn results in higher tool wear. Simultaneously the maximum effective plastic strain is reduced due to the reduced ability of the material to flow freely in the die cavity. The findings confirm that friction is one of the most important technological parameters affecting the stability, efficiency and the general quality of the hot forging process.

Future research may include a more extensive parametric analysis involving different process conditions, as well as the validation of numerical results by comparison with experimental and industrial production data. Such an approach would further improve the accuracy of the simulation model and contribute to the development of more efficient forging technologies.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Zastava Kovačnica A.D., Kragujevac, Serbia, for providing the technical resources, engineering infrastructure, and support necessary for carrying out the numerical simulations and research presented in this paper.

References

[1] LEE, R.S., JOU, J.L.: *Application of numerical simulation for wear analysis of warm forging die*, Journal of Materials Processing Technology, 140, 2003.

[2] HARTLEY, P., PILLINGER, I.: *Numerical simulation of the forging process*, Comput. Methods Appl. Mech. Engrg., 195, 2006.

[3] DELIĆ, M., DELIĆ, M.: *Analysis of the influence of hot forging process parameters on tool wear using the finite element method,* Faculty of Engineering, Kragujevac, 2025.

[4] HAWRYLUK, M., GRONOSTAJSKI, Z., ZIEMBA, J., DWORZAK, Ł., JABŁOŃSKI, P., RYCHLIK, M.: *Analysis of the influence of lubrication conditions on tool wear used in hot die forging processes*, Eksploatacja i Niezawodnosc - Maintenance and Reliability, 20, 2018.

[5] DAVOUDI, M., NEJAD, A.F., KOLOOR, S.S.R., PETRŮ, M.: *Investigation of effective geometrical parameters on wear of hot forging die*, Journal of Materials Research and Technology, 15, 2021.

[6] YIN, J., HU, R., SHU, X.: *Closed-die forging process of copper alloy valve body: finite element simulation and experiments*, Journal of Materials Research and Technology, 10, 2021.

[7] ŘIHÁČEK, J., PETERKOVA, E., CISAROVA, M., KUBICEK, J.: *Comparison of FEM and FVM for the numerical simulation of forging process*, MM Science Journal, 2020.

[8] CHENOT, J.-L., BOUCHARD, P.-O., FOURMENT, L., LASNE, P.: *Numerical Simulation and Optimization of the Forging Process*, International Cold Forging Congress 12th ICFC, Stuttgart, 2011.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CAE SIMULATION OF THE HOT FORGING PROCESS FOR A MILITARY VEHICLE STEEL COMPONENT

MARKO DELIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; marko.delic@kg.ac.rs

NIKOLA KASTRATOVIĆ
Zastava Kovačnica A.D. Kragujevac, Kosovska 4b, 34000 Kragujevac, Serbia; nikolakastratovic93@gmail.com

VESNA MANDIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; mandic@kg.ac.rs

DUŠAN ARSIĆ
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; dusan.arsic@fink.rs

NIKOLA KOMATINA
Faculty of Engineering University of Kragujevac,Sestre Janjić 6, 34000 Kragujevac, Serbia; nkomatina@kg.ac.rs

Abstract: *Today, it is impossible to imagine mechanical parts in the automotive, mechanical or military industry without forgings, primarily because of their durability and safety in operation. However, in order to obtain a quality part, one must know exactly how the material flows through the tool and what happens during the deformation itself. That's why numerical simulations have become our main tool when developing or changing forging technology.*

Instead of expensive tests in the plant, we used a computer model to follow in detail the propagation of deformations, the change of stress and the movement of the forging force during the compression of the steel. We mainly focused on identifying critical regions in the tool that can adversely affect the geometry of the forging or cause unexpected production downtime.

The results from Simufact gave us in practice a "window" into the interior of the material during compaction. Based on that, we precisely assessed whether the proposed technological solution is good, monitoring how the mold is filled and how efficient the whole process actually is.
This engineering approach through CAE methods clearly proves its value in the industry, because it directly saves us time for new product development.

Keywords: *forging, Simufact Forming, material flow, military vehicle, CAE*

1. INTRODUCTION

The hot forging is one of the most important technologies for metal forming. New technologies like 3D metal printing have been developed, but hot forging is still very important in the production process of parts under high loads. Due to good mechanical properties of manufactured parts, hot forging has an important role on military and automotive industries. Lower values of the yield stress are obtained when the workpiece is heated up. But it is very difficult to lubricate because of the high temperature. In hot forging process, the tools are subjected to high temperature, friction due to sliding of material on tool surface and in case of hammer forging, impact load.

The ever-increasing costs of material, energy and especially manpower demand that forging processes and tooling be designed and developed with minimum amount of trial and error with shortest possible lead times. Hence, to keep up with the competition the cost-effective application of computer aided techniques i.e. CAD, CAM, CAE and especially Finite Element Analysis (FEA) and Finite Volume Analysis (FVA) based computer simulation are an absolute necessity. The practical use of these techniques requires knowledge of the principal variables of the forging process and their interaction.

CAE simulations of processes using software based on the finite element method are successfully applied to eliminate material flow defects, determine and optimize process parameters, determine tool stresses to prevent tool breakage, and predict workpiece properties through stress-strain analysis. In addition, it is very easy to make changes to the tool design and monitor the impact of these changes on process parameters, thus achieving optimal tool geometry at all stages of the process.

In hot forging processes, the lifetime of forging tools

Corresponding author: N. Kastratović;
E-mail: nikolakastratovic93@gmail.com
DOI: 10.5937/OTEH26134D

(upper and lower dies) is largely limited by the occurrence of abrasive wear, thermal fatigue and plastic deformation. Research shows that wear is the dominant failure mechanism, responsible for as much as 70% of all damage to tools. Validation of numerical tool wear models is performed by comparing simulation results with experimental measurements on three-dimensional measuring machines (CMM) [1].

Lee and Jou [2] applied Archard's wear law within a finite volume framework to estimate the service life of hot forging tools. Their approach enabled the prediction of wear development based on local contact conditions such as pressure, sliding velocity and temperature.

Hartley and Pillinger [3] highlighted the importance of remeshing techniques in forging simulations due to severe element distortion caused by large plastic deformations. Their findings demonstrated that adaptive mesh refinement is essential for maintaining numerical stability.

Hawryluk et al. [4] investigated tribological conditions in hot forging processes and showed that friction conditions continuously evolve as a consequence of lubricant degradation and elevated temperatures.

Davoudi et al. [5] emphasized the importance of accurate contact stress prediction in forging tool analysis. Their work demonstrated that fine mesh discretization improves the identification of stress concentration zones and increases prediction accuracy.

Yin et al. [6] reported that improper billet positioning may result in eccentric loading and premature tool failure, highlighting the importance of accurate billet alignment during forging operations.

Řiháček et al. [7] compared different numerical approaches implemented in Simufact Forming and concluded that the finite volume method provides superior numerical stability in simulations involving severe plastic deformation and intensive material flow.

Chenot et al. [8] demonstrated that numerical simulation can significantly reduce the need for costly experimental trials by enabling process optimization in a virtual environment before industrial implementation.

Studies on hot deformation behavior of Cr-alloyed steels have shown that dynamic recrystallization is the dominant softening mechanism at elevated temperatures. The flow stress response is strongly dependent on temperature and strain rate, making accurate material modelling essential for reliable numerical simulations [9].

2. SIMULATION CONDITIONS

A three-dimensional thermo-mechanical finite element model was developed using Simufact Forming software in order to analysis the upset forging process of an industrial component manufactured from 38HS steel.

Steel 38HS (according to the standard GOST 4543) is an improved steel which is highly valued in the forging industry due to its high strength, excellent toughness and good hardenability. The material has a precise balance of chromium and molybdenum, which gives a superior resistance to material fatigue and impact loads and is the ideal choice for the manufacture of components that are highly stressed in the defense and automotive industries. 38HS exhibits a stable behavior at high temperatures during the hot forging process, enabling the production of complex geometries without the risk of internal defects. After a suitable thermal treatment, 38HS shows superior mechanical properties required to operate in the most challenging working conditions.

The simulation model consisted of three main geometrical entities: the workpiece, the upper die and the lower die. The workpiece geometry was imported as a CAD model and discretized using an adaptive finite element mesh.

The workpiece material was defined as 38HS. The material behavior was considered temperature-dependent in order to accurately represent the hot forging conditions.

The forging process was simulated through three consecutive stages: upsetting, pre-forging and final forging. The upsetting and pre-forging operations were analyzed using a single mesh configuration, since their primary purpose was to generate the intermediate billet geometry required for the final forging stage.

A mesh sensitivity analysis was performed exclusively for the final forging operation, which represents the most critical stage of the process due to the complex die geometry and intensive material flow. Three mesh sizes were investigated: 3.0 mm, 2.2 mm and 1.5 mm. All other process parameters, including billet temperature, die temperature, friction factor and press characteristics, were kept constant throughout the analysis.

This approach enabled the influence of mesh density on numerical stability, computational time and predicted process parameters to be evaluated while maintaining identical initial conditions obtained from the upsetting and pre-forging simulations.

The simulations were performed using the finite volume (FV) solver in Simufact Forming. The initial billet temperature was set to 1250 °C, while the die temperature was maintained at 250 °C. A shear friction model with a friction factor of m = 0.5 was applied at the workpiece–die interface. The forging operation was simulated using a Forging hammer with an impact energy of 35 kJ and an efficiency factor of 0.7. The total forging stroke was 70 mm and each simulation was divided into 100 equal increments.

Therefore, the only variable parameter among the analyzed cases was the mesh density, allowing the determination of the influence of element size on numerical stability, computational time and predicted process parameters to be evaluated.

3. RESULTS AND DISCUSSION

The numerical simulation of the hot forging process was carried out using the finite volume solver implemented in Simufact Forming. The objective of the analysis was to evaluate material flow, strain distribution, contact conditions and the influence of mesh density on the numerical stability of the obtained results.

The simulation results confirmed complete filling of the die cavity and successful material flow throughout the forging operation. The highest values of effective plastic strain were recorded in the flash region and in areas characterized by intensive material redistribution. Such behavior is expected in hot forging processes due to the constrained flow conditions near the die parting line.

The computational cost increased significantly with mesh refinement. The simulation time increased from approximately 34 minutes for the 3.0 mm mesh to 90 minutes for the 2.2 mm mesh and 111 minutes for the 1.5 mm mesh. In addition, the number of remeshing operations increased from four to six, indicating a higher degree of mesh distortion during the forming process.

Because similar anomalies appeared in both refined meshes, it can be concluded that the observed behavior is related to local numerical effects rather than to the actual physical response of the material. Consequently, the 3.0 mm mesh provided the most stable and physically meaningful solution while maintaining acceptable computational efficiency.

Table 1 presents the influence of mesh density on the maximum effective plastic strain and maximum contact pressure obtained during the final forging simulation. The results indicate that mesh refinement leads to a gradual increase in both parameters. The maximum effective plastic strain increased from 1.92 for the 3.0 mm mesh to 2.07 and 2.09 for the 2.2 mm and 1.5 mm meshes, respectively. The relatively small difference between the 2.2 mm and 1.5 mm meshes suggests that the solution approaches numerical convergence with respect to plastic strain.

A similar trend was observed for the maximum contact pressure, which increased from 1215 MPa for the 3.0 mm mesh to 1499 MPa and 1790 MPa for the 2.2 mm and 1.5 mm meshes, respectively. This behavior is expected because finer meshes provide a more accurate representation of local contact conditions and stress concentrations, particularly in regions with complex die geometry.

The obtained results demonstrate that CAE-based process simulation is an effective tool for evaluating forging technology before industrial implementation. The performed analyses enabled detailed insight into material flow, strain development and die filling conditions, reducing the need for costly experimental trials and shortening the development cycle of forged components intended for military vehicle applications.

Figure 1 shows a 3D model of the tool assembly, which includes the upper and lower dies imported in the Simufact Forming software environment. The geometry of the tool is imported from the CAD system with the application of precise discretization, which ensures a reliable mapping of contact surfaces and interactions with the workpiece during the forging process. Figure 1 also show forged part.

Table 1. Comparation of results

Parameter	Mesh 3.0 mm	Mesh 2.2 mm	Mesh 1.5 mm
Max effective plastic strain	1.92	2.07	2.09
Max pressure (MPa)	1215	1499	1790
Max strain rate (s^{-1})	173982	402216	364102
Max die wear (mm)	1.67E-11	2.53E-11	2.06E-11
Max temperature (°C)	1351.52	1372.61	1537.31

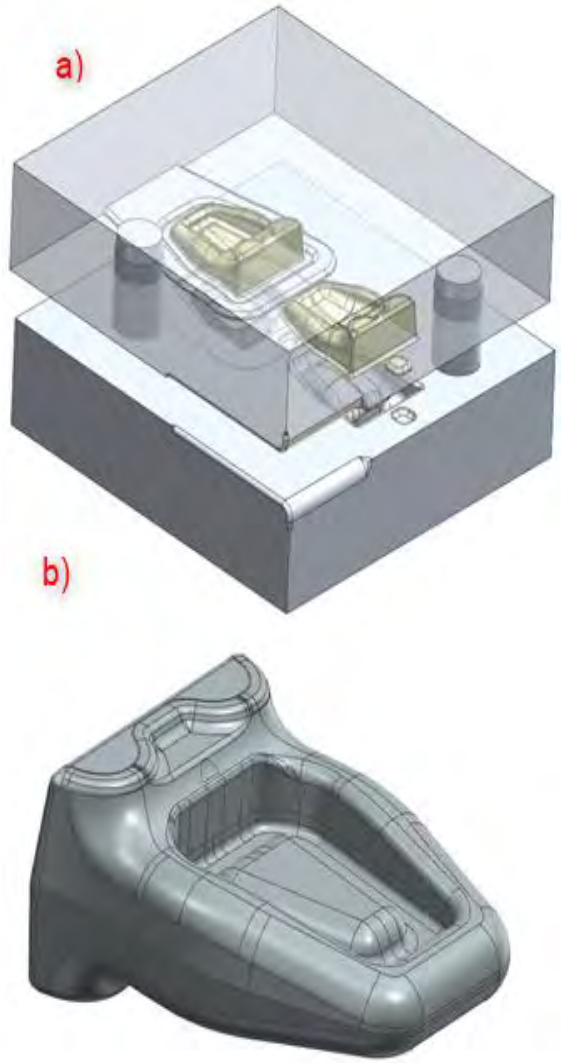


Figure 1. 3D model: a) Upper and lower die for forging; b) Forged part

Figure 2 presents the upsetting and pre-forging stages used to obtain the intermediate billet geometry for the final forging operation. These stages were identical for all analyzed cases and were simulated using the same process parameters and mesh configuration.

Figures 3, 4 and 5 illustrate the final forging stage

performed with different mesh densities in order to evaluate the influence of element size on the numerical results. Figure 3 shows the results obtained using a 3.0 mm mesh, Figure 4 presents the simulation performed with a 2.2 mm mesh, while Figure 5 corresponds to the finest analyzed mesh of 1.5 mm.

All simulations were carried out under identical technological conditions, including the initial billet temperature of 1250 °C, die temperature of 250 °C, friction factor of μ = 0.5 and hammer energy of 35 kJ. Therefore, the only varying parameter among the analyzed cases was the mesh density used in the final forging stage.

The comparison of the obtained results enabled the evaluation of mesh sensitivity, computational efficiency and numerical stability, while maintaining identical initial conditions generated during the upsetting and pre-forging operations.

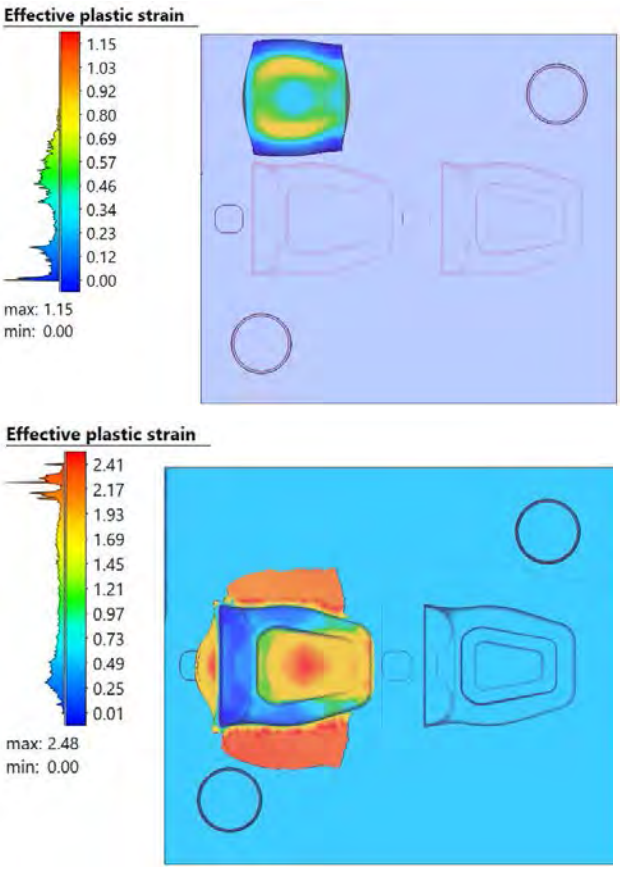


Figure 2. Upsetting and pre-forging

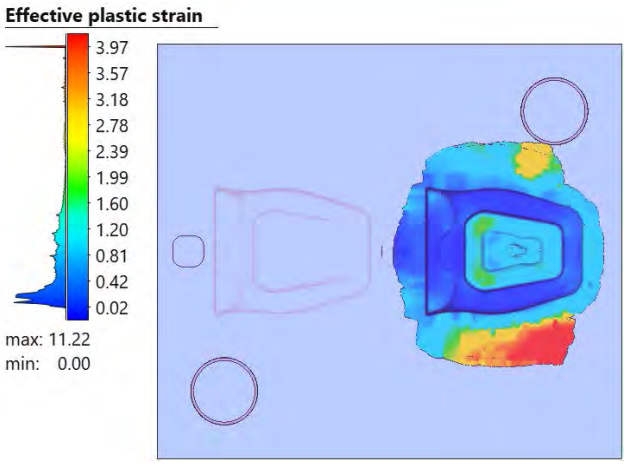


Figure 3. Value of effective plastic strain mesh 3 mm

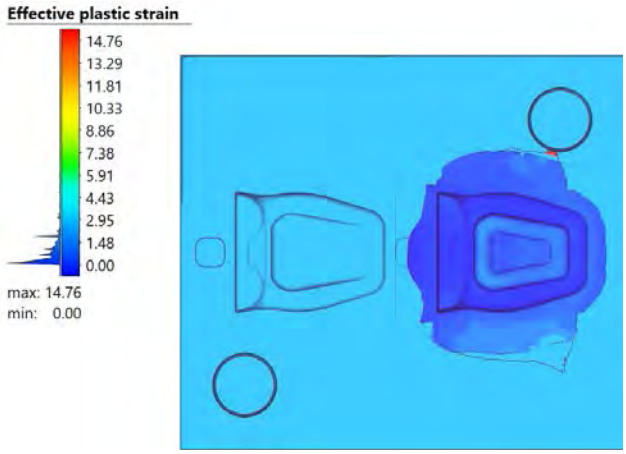


Figure 4. Value of effective plastic strain mesh 2.2 mm

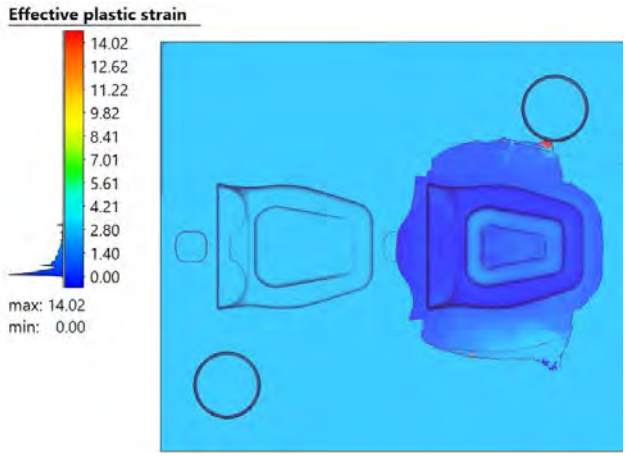


Figure 5. Value of effective plastic strain mesh 1.5 mm

Figure 6 shows the distribution of effective plastic strain for the three mesh configurations investigated. In all cases, the highest values are observed in the forgings wreath region, where the material is subjected to severe compression as it flows into the narrow flash gap. These localized maxima are expected and represent the most intensive plastic deformation occurring during the forging process rather than the deformation of the forged component itself. Therefore, the effective plastic strain values reported in Table 1 were measured directly on the forged part, excluding the forgings wreath region, in order

to provide a representative comparison of the deformation state within the final product. The results indicate only minor differences between the investigated mesh sizes, confirming that mesh refinement has a limited influence on the predicted deformation of the forged component, while mainly affecting the resolution of localized deformation in the forgings wreath region.

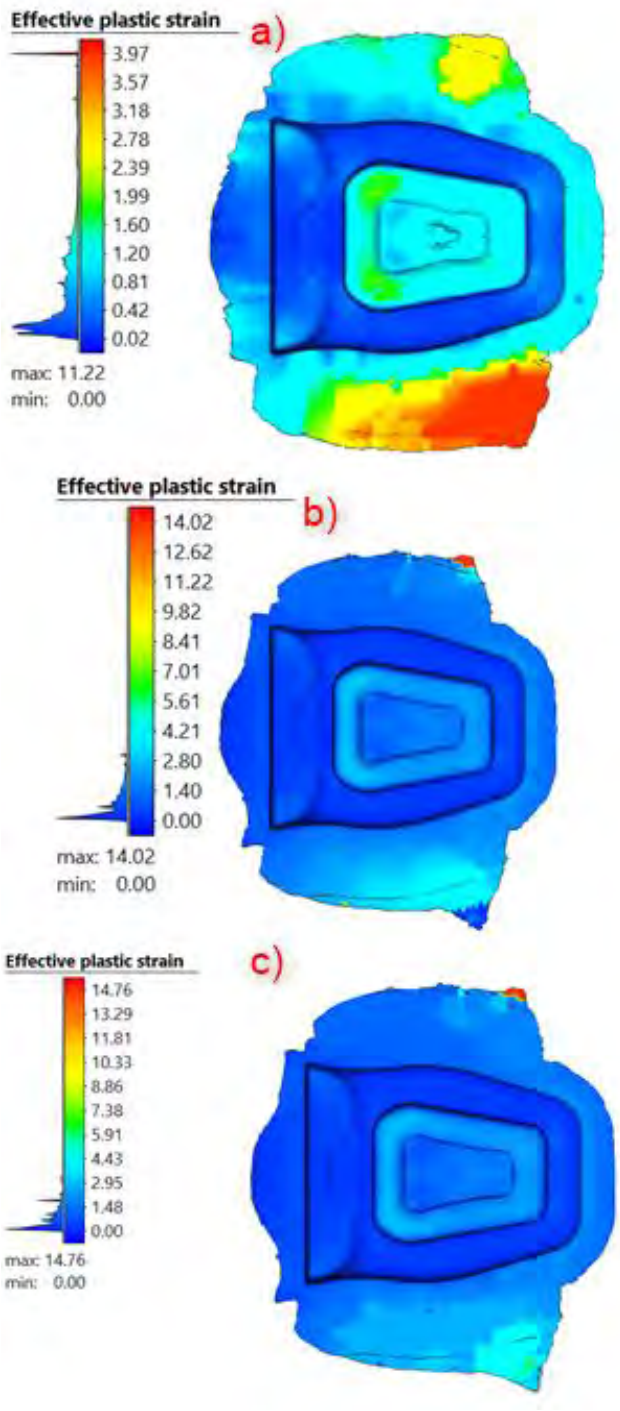


Figure 6. Forged part: a) Mesh 3.0 mm b) Mesh 2.2 mm c) Mesh 1.5 mm

5. CONCLUSION

In this paper, a numerical analysis of the forging process was carried out, where the finite volume method (FVM) was applied and showed exceptional stability and reliability in the simulation of large plastic deformations.

This study presented a numerical simulation of the hot forging process of a 38HS steel component intended for military vehicle applications using Simufact Forming software. The conducted CAE analysis enabled a detailed investigation of material flow, strain distribution, temperature evolution and die filling behavior throughout the forging operation.

The simulation results confirmed complete die cavity filling and demonstrated that the proposed forging technology is capable of producing the required component geometry without critical defects that could compromise the manufacturing process. The highest deformation values were observed in the flash region, which is consistent with the expected material flow characteristics of hot forging processes.

A mesh sensitivity analysis was performed using element sizes of 3.0 mm, 2.2 mm and 1.5 mm. Although finer meshes provided a more detailed representation of local deformation phenomena, they significantly increased computational time and introduced numerical instabilities manifested through nonphysical pressure, density and temperature peaks. Among the analyzed models, the 3.0 mm mesh provided the most stable and computationally efficient solution while maintaining physically meaningful results.

The obtained findings demonstrate that CAE-based simulation represents a valuable engineering tool for the design and optimization of forging technologies. By enabling virtual process verification prior to industrial implementation, numerical simulation reduces development costs, minimizes experimental trials and shortens product development time.

The methodology applied in this study can be successfully extended to other forged components manufactured from 38HS steel and may serve as a basis for future research focused on process optimization, tool design improvement and advanced thermo-mechanical analysis of hot forging operations.

Future work may involve the investigation of additional process parameters and their impact on forging performance. The incorporation of experimental validation and industrial production data would allow for further calibration of the numerical model, leading to improved prediction accuracy and enhanced process optimization in industrial forging applications.

Acknowledgements

The authors gratefully acknowledge Zastava Kovačnica A.D., Kragujevac, Serbia, for providing the technical resources, engineering infrastructure, and support necessary for carrying out the numerical simulations and research presented in this paper.

References

[1] NARESH, P., BABU, A. V. H., MADHAVA, V., REDDY, M. S.: *Design and Analysis of a Hot Forging Dies*, Journal of Mechanics and MEMS

(JMM), 2016.

[2] LEE, R.S., JOU, J.L.: *Application of numerical simulation for wear analysis of warm forging die*, Journal of Materials Processing Technology, 140, 2003.

[3] HARTLEY, P., PILLINGER, I.: *Numerical simulation of the forging process*, Comput. Methods Appl. Mech. Engrg., 195, 2006.

[4] HAWRYLUK, M., GRONOSTAJSKI, Z., ZIEMBA, J., DWORZAK, Ł., JABŁOŃSKI, P., RYCHLIK, M.: *Analysis of the influence of lubrication conditions on tool wear used in hot die forging processes*, Eksploatacja i Niezawodnosc - Maintenance and Reliability, 20, 2018.

[5] DAVOUDI, M., NEJAD, A.F., KOLOOR, S.S.R., PETRŮ, M.: *Investigation of effective geometrical parameters on wear of hot forging die*, Journal of Materials Research and Technology, 15, 2021.

[6] YIN, J., HU, R., SHU, X.: *Closed-die forging process of copper alloy valve body: finite element simulation and experiments*, Journal of Materials Research and Technology, 10, 2021.

[7] ŘIHÁČEK, J., PETERKOVA, E., CISAROVA, M., KUBICEK, J.: *Comparison of FEM and FVM for the numerical simulation of forging process*, MM Science Journal, 2020.

[8] CHENOT, J.-L., BOUCHARD, P.-O., FOURMENT, L., LASNE, P.: *Numerical Simulation and Optimization of the Forging Process*, International Cold Forging Congress 12th ICFC, Stuttgart, 2011.

[9] MENGARONI, S., CIANETTI, F., CALDERINI, M., EVANGELISTA, E., DI SCHINO, A., MCQUEEN, H.: *Tool Steels: Forging Simulation and Microstructure Evolution of Large Scale Ingot*, Acta Physica Polonica A, 2015.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

OPTIMIZATION OF INTERFACIAL BONDING OF EPDM-BASED THERMAL INSULATION AND STEEL ROCKET MOTOR CASE-DEVELOPMENT OF THE SYSTEM OF ADHESIVE LAYER AND CEMENT

OLGA MARAŠ
Military Technical Institute, Belgrade, olgamaras98@gmail.com

SARA PEPIĆ
Military Technical Institute, Belgrade, spepic00@gmail.com

JELENA GRŽETIĆ
Military Technical Institute, Belgrade, jrusmirovic@tmf.bg.ac.rs

PETAR VASILJEVIĆ
Military Technical Institute, Belgrade, peki.v94@gmail.com

Abstract: *The design of rocket motors with case-bonded propellant grains requires reliable thermal protection systems capable of protecting the motor case from high thermal and mechanical loads during operation. Due to that, it is important to apply thermally stable insulating material using a suitable technology and achieve good adhesion between thermal insulation and rocket motor case. This paper presents the development of the system of adhesive layer and cement for optimal interfacial bonding of EPDM based thermal insulation and steel rocket motor case. An EPDM thermal insulation compound reinforced with chopped and milled carbon fibers was selected for application to the rocket motor case and further evaluated through thermal analysis and adhesion testing. The gluing of EPDM-based thermal insulation to the interior of the rocket motor case was performed by vulcanization in an autoclave using the adhesive layer and cement as adhesion promoters. Thermal stability of the insulation material was evaluated by simultaneous thermal analysis (STA), while adhesion between rubber-rubber and rubber-steel was determined according to ASTM D1876 and ASTM D6862 standards. The results confirmed satisfactory adhesion properties of the selected adhesive system and adequate thermal stability of the insulation compound. The rocket motor case was successfully lined with the developed EPDM-based thermal insulation in a thickness range of 1.8–2.2 mm and vulcanized under industrial processing conditions.*

Keywords: *EPDM rubber, thermal insulation rocket motor case, adhesion, thermal protection*

1. INTRODUCTION

Rocket motors with case-bonded propellant grains require thermal protection systems capable of protecting the motor case from the extreme thermal and mechanical loads generated during motor operation . Thermal insulation is applied directly to the inner wall of the motor case to reduce heat transfer from combustion gases and prevent overheating of the metallic structure. In addition to thermal protection, the insulation layer may also serve as an inhibitor, protecting selected propellant surfaces during combustion [1, 2, 3].

Elastomer-based thermal insulation materials are widely used in rocket motor applications due to their processability, elasticity and thermal stability. Among them, EPDM rubber is of particular interest because of its resistance to thermal ageing and oxidation. In the investigated thermal insulation compound, chopped and milled carbon fibers, ammonium polyphosphate and melamine were used to improve thermal protection performance [4, 5, 6].

A critical requirement for the reliable operation of case-bonded rocket motor systems is the interfacial adhesion between the thermal insulation material and the metal motor case. Insufficient adhesion may lead to delamination, local heat accumulation and degradation of structural integrity during service [7]. For this reason, the development of the system of adhesive layer and cement and the suitable vulcanization technology must ensure an optimal interfacial bonding between thermal insulation and the metal rocket motor case.

The aim of this work was to optimize the interfacial bonding of EPDM-based thermal insulation through the development of a system of adhesive layer and cement and the development of suitable vulcanization technology for applying an EPDM-based thermal insulation to the inner wall of a steel rocket motor case.

2. EXPERIMENTAL

An EPDM-based thermal was used for application to the inner wall of a steel rocket motor case. The thermal

Corresponding author: O. Maraš;
E-mail: olgamaras98@gmail.com
DOI: 10.5937/OTEH26135M

insulation was designated TG-H-727/3. The composition of the investigated thermal insulation compound is presented in Table 1. A sulfur-based curing system consisting of mercaptobenzothiazole (MBT), tetramethylthiuram disulfide (TMTD), and sulfur was used for vulcanization of the experimental compounds. Commercially available processing additives were also incorporated at optimum levels to improve processability. The thermal insulation was designated TG-H-727/3. Homogenization of the compounds was performed on a two-roll mill. After mixing, test specimens were prepared and curing conditions were determined from rheological measurements, while thermal stability was evaluated by STA.

Table 1. Formulation of EPDM-based thermal insulation (TI) compound

Components for TI compound formulation (TG-H-727/3)	Parts per hundred rubber / Mass percentage (phr/%)
EPDM rubber Dutral TER 9046	100.0 /42.9
Silicon dioxide, SiO2 VULKASIL S	15.00/6.4
Semi-reinforcing carbon black N550	15.0/6.4
Melamine	35.0 /15.0
Ammonium polyphosphate (APP)	35.0 /15.0
Carbon fibers, milled 0.1 mm	5.0/2.1
Carbon fibers, chopped 3.0 mm	5.0 /2.1
Chlorinated paraffin oil	3.5 /1.5
Paraffinic oil RPO320	10.0 /4.3
Zinc oxide, ZnO	5.0 /2.1
Stearic acid	1.0 /0.4
Sulfur	1.5 /0.6
MBT – Mercaptobenzothiazole	1.5 /0.6
TMTD – Tetramethylthiuram disulfide	0.5 /0.2
Total:	233.0 /100.0

An adhesive layer designated TG-HD-727/C3 and a cement designated TG-D-727/C6 were developed to provide bonding between the thermal insulation layer and the metal substrate. The compositions of the adhesive layer and cement are presented in Tables 2, 3 and 4, respectively. Adhesion properties were determined using peel and shear tests according to ASTM D1876 and ASTM D6862.

Table 2. Formulation of the adhesive layer compound

Components for adhesive layer formulation (TG-HD-727/C3)	Parts per hundred rubber (phr/%)
EPDM rubber Keltan 6950 C (or Dutral TER 9046)	60.0/28.2
Polychloroprene rubber, CR rubber, Baypren 110	40.0/18.8
Silicon dioxide, SiO2 VULKASIL S	15.0/7.1
Melamine	35.0/16.5
Ammonium polyphosphate (APP)	35.0/16.5
Zinc oxide, ZnO	5.0/2.4
Stearine	1.0/0.5
Magnesium oxide, MgO	1.0/0.5
Chlorinated paraffin oil	10.0/4.7
Adhesion promoter, Elastobond T 6000	2.0/0.9
Adhesion promoter, Elastobond T 6000	5.0/2.3
Sulfur	1.5/0.7
MBT–Mercaptobenzothiazole	1.5/0.7
TMTD–Tetramethylthiuram disulfide	0.5/0.2
Total:	212.5/100.0

Table 3. Composition of TG-D-727/C6 cement

Components	Parts per hundred rubber (phr/%)
CR rubber, Baypren 110	100.0/63.7
Silicon dioxide, SiO2 VULKASIL S	35.0/22.3
Plasticizer, Pacizer 6218	5.0/3.2
Zinc oxide, ZnO	12.0/7.6
Stearine	1.0/0.6
Accelerator, Rhenogran ETU-80	1.0/0.6
Magnesium oxide, MgO	1.0/0.6
Antioxidant / Anti-aging agent, Vulkanox SKF	2.0/1.3
Total:	157.0/100.0

Table 4. Composition of organic solvents and other components for TG-D-727/C6 rubber cement preparation

Components	Mass (g)
TG-D-727/C6	75.0
Petroleum ether / Gasoline	90.0
Toluene	90.0
Acetone	70.0
Resin, Koresin PLW	5.9
Adhesion promoter, Elastobond T 6000	4.6

The rocket motor case made of Č.4734 steel (according to the YUS) was prepared by abrasive blasting of the inner surface, followed by application of the system of adhesive layer and cement. The thermal insulation lining consisted of an adhesive layer with a thickness of 0.6 mm and a thermal insulation layer with a thickness from 1.4 mm to 1.6 mm, resulting in a total lining thickness between 1.8 and 2.2 mm.

After application, the lined rocket motor cases were vulcanized in an autoclave using direct steam at 150 °C and a pressure of 0.5 MPa for 40 min.

3. RESULTS AND DISCUSSION

3.1. Thermal Stability of the EPDM-based Thermal Insulation

The thermal stability of the investigated EPDM-based thermal insulation was evaluated by STA. The TG curve (Figure 1) showed a negligible mass loss below 250 °C, which can be attributed to the evaporation of residual low-molecular-weight components and moisture. The main degradation stage occurred between approximately 250 °C and 480 °C and corresponded to the thermal decomposition of the EPDM matrix and the release of volatile pyrolysis products. The compound exhibited a final mass loss of 71.87 %, while maintaining sufficient thermal stability throughout the temperature range relevant for the investigated application.

The DTG curve (Figure 2) revealed three characteristic degradation peaks located at approximately 298 °C, 384 °C and 464 °C. The first peak was associated with the decomposition of melamine and ammonium polyphosphate, the second with the formation of a protective carbonized layer, and the third with the primary degradation of the EPDM matrix. The obtained results confirmed the thermal behavior of the selected insulation compound and supported its use in the thermal protection system of the rocket motor case [8].

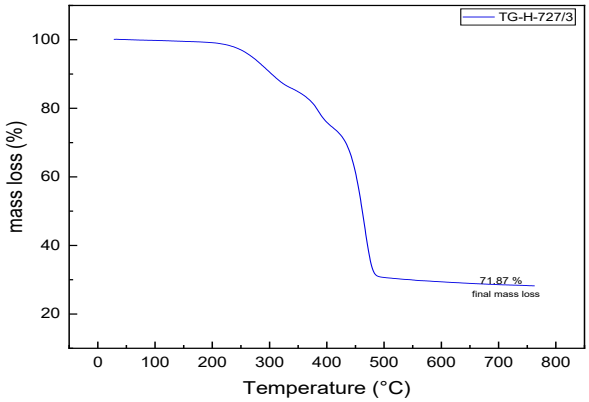


Figure 1. TG curve of the TG-H-727/3 formulations

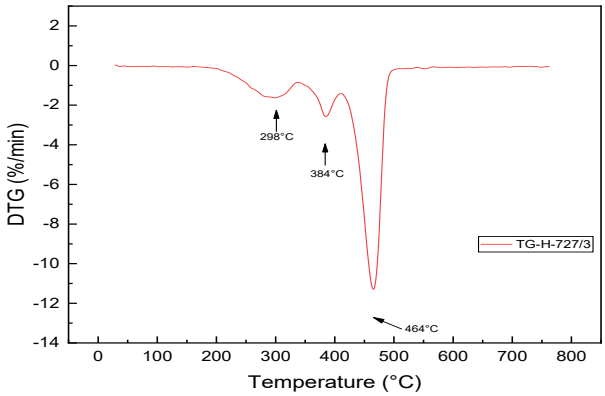


Figure 2. DTG curve of the TG-H-727/3 formulations

3.2. Adhesion Properties and Verification of the Application Technology

The sample codes used in the adhesion tests are defined as follows: T denotes the EPDM-based thermal insulation compound TG-H-727/3; P denotes the adhesive layer based on TG-HD-727/C3; C3 denotes TG-HD-727/C3 when used as a cement; C6 denotes the TG-D-727/C6 cement. For comparative rubber-rubber and rubber-metal adhesion testing, an adhesive based on natural rubber (NR), designated MC, was also used. The number in the sample code identifies the individual specimen [8].

The rubber-rubber peel test results (Table 5) revealed a strong influence of both the adhesive layer and the selected cement on the bond strength. The highest average peel strength (10.63 N/cm) was obtained for specimen T/P-C3-C3-T/P-C3, which contained the TG-HD-727/C3 adhesive layer and the C3 cement on both bonded surfaces. Failure occurred within the adhesive layer, indicating good compatibility between the thermal insulation, the adhesive layer and the cement system.
Specimens prepared without the additional adhesive-layer application exhibited significantly lower peel strength values. The average peel strength of specimen T-C3-T was 4.01 N/cm, while specimen T-C6-T showed the lowest value of 2.82 N/cm. These results demonstrate that the incorporation of the TG-HD-727/C3 adhesive layer considerably improves adhesion between thermal insulation layers. The intermediate value obtained for

specimen T/P-C6-C6-T/P-C6 (5.30 N/cm) suggests that the adhesive layer alone is not sufficient to achieve maximum bond strength and that the choice of cement also plays an important role.

Table 5. Rubber-rubber peel strength of the EPDM-based thermal insulation system

sample code	L (mm)	F_{max} (N)	F_{min} (N)	F_{avg} (N)	$F_{avg/L}$ (N/cm)	mean value (N/cm)
1-1 (T/P-C3-C6-T/P-C3)	31.65	27.00	24.00	25.32	8.0	6.70
1-2 (T/P-C3-C6-T/P-C3)	30.96	17.60	15.50	16.64	5.39	
2-1 (T-C6-T)	32.45	13.70	11.20	12.50	3.85	2.82
2-2 (T-C6-T)	31.28	7.10	4.00	5.63	1.80	
3-1 (T/P-C3-C3-T/P-C3)	30.83	27.88	26.60	27.70	9.04	10.63
3-2 (T/P-C3-C3-T/P-C3)	31.40	38.70	36.60	38.38	12.22	
4-1 (T-C3-T)	31.13	9.00	5.70	7.90	2.54	4.01
4-2 (T-C3-T)	32.04	18.70	16.80	17.56	5.48	
5-1 (T/P-C6-C6-T/P-C6)	31.67	21.00	14.50	16.79	5.30	5.30

The rubber–metal peel test results (Table 6) showed significant differences in adhesion performance among the investigated bonding systems. Specimens T-C6 (1-1), T-C6 (1-2), T-C3 (2-1) and T-C3 (2-2) represent the bonding systems based on C6 and C3, while T-MC represents the comparative MC adhesive. The highest peel strength was obtained for specimen T-MC, with an average value of 15.99 N/cm. Among the investigated C3 and C6 systems, specimen T-C6 (1-1) exhibited the highest peel strength of 5.03 N/cm. The results demonstrate differences in the adhesion performance of the investigated bonding systems for joining the EPDM-based thermal insulation to the metal substrate.

Table 6. Rubber-metal peel strength of the EPDM-based thermal insulation system

sample code	L (mm)	F_{max} (N)	F_{avg} (N)	$F_{avg/L}$ (N/cm)
1-1 T-C6	32.55	18.40	16.37	5.03
1-2 T-C6	32.29	5.40	4.10	1.27
2-1 T-C3	32.68	12.90	12.43	3.80
2-2 T-C3	32.38	11.10	9.07	2.80
T-MC	31.84	49.90	9.64	15.99

The results of the rubber-rubber shear test (Table 7) showed that specimens T-C3-T (2-1) and T-C3-T (2-2) exhibited the highest shear stresses, reaching 0.232 N/mm² and 0.207 N/mm², respectively. During testing, elongation and tearing of the rubber were observed, indicating good adhesion and high mechanical resistance of the bonded joint. In contrast, specimens T-MC-T (3-1) and T-MC-T (3-2) exhibited lower shear stresses, accompanied by interfacial debonding, which indicates weaker adhesion compared with the systems based on C3 and C6 cements.

Table 7. Rubber-rubber shear strength of the EPDM-based thermal insulation system

sample code	L (mm)	b (mm)	A (mm^2)	F_{max} (N)	F_{avg} (N)	F_{max}/A (N/mm^2)
1-1 T-C6-T	21.35	21	448.35	53.25	53.1	0.119
1-2 T-C6-T	21.71	21	455.91	84.9	80.94	0.185
2-1 T-C3-T	22.30	21	468.3	108.8	108.01	0.232
2-2 T-C3-T	22.35	21	469.35	97.10	95.70	0.207
3-1 T-MC-T	21.41	21	449.61	76.8	72.61	0.171
3-2 T-MC-T	21.43	21	450.03	68.89	68.41	0.153

The rubber-metal shear test results (Table 8) showed similar shear stress values for specimens T-C6 and T-C3, amounting to 0.287 N/mm² and 0.240 N/mm², respectively. These results indicate good adhesion between the EPDM-based thermal insulation system and the steel substrate. For specimen T-C6, failure occurred through rubber tearing rather than debonding at the interface, suggesting that the adhesive bond strength exceeded the cohesive strength of the rubber.

Table 8. Rubber-metal shear strength of the EPDM-based thermal insulation system

sample code	L (mm)	b (mm)	A (mm^2)	F_{max} (N)	F_{avg} (N)	$F_{max/A}$ (N/mm^2)
T-C6	21.83	21	458.43	131.5	103.47	0.287
T-C3	21.31	21	447.51	107.3	105.18	0.240

Both specimens, T-MC-T (1-1) and T-MC-T (1-2), exhibited high shear stress values (Table 9). The slightly higher value obtained for specimen T-MC-T (1-1) indicates good homogeneity of the bonded region and efficient crosslinking between the two rubber layers. The obtained results confirm the ability of the MC adhesive to produce mechanically stable joints between the rubber layers.

Table 9. Shear strength of bonded rubber layers in the EPDM-based thermal insulation system

sample code	L (mm)	b (mm)	A (mm^2)	F_{max} (N)	F_{sr} (N)	F_{max}/A (N/mm^2)	
1-1 T-MC-T	27.12	3.35	90.85	75.30	75.30	0.829	0.793
1-2 T-MC-T	25.83	3.35	86.53	65.60	65.50	0.758	

3.3. Application and Dimensional Verification of the Thermal Insulation Lining

The developed EPDM-based thermal insulation system was successfully applied to the inner surface of the steel motor case using the selected adhesive layer TG-HD-727/C3, TG-D-727/C6 cement and thermal insulation compound TG-H-727/3. The application process included surface preparation by sandblasting, adhesive treatment of the chamber wall, positioning of the insulation layer (Figure 3a), air-assisted blowing (Figure 3b), and subsequent vulcanization in an autoclave at 150 °C and 0.5 MPa for 40 min [8].

Particular attention was paid to achieving a uniform insulation thickness and reliable bonding between the insulation layer and the metallic substrate. The use of a duplicated insulation structure consisting of a 0.6 mm adhesive layer and a 1.4 mm to 1.6 mm thermal insulation layer enabled stable positioning of the material and prevented displacement during processing. The selected joining procedure, including a 45° scarf joint and pressure-assisted consolidation, ensured continuous contact between adjacent insulation segments and minimized the possibility of local defects or air entrapment.

Figure 3a.

Figure 3b.

Figures 3 (a and b) Application process of the EPDM-based thermal insulation lining to the rocket motor case

Dimensional verification performed after application of the adhesive layer and thermal insulation confirmed that the final coating thickness was within the target range of 1.8–2.2 mm throughout the chamber length. The measured inner diameter values demonstrated good process repeatability and indicated that the insulation layer was distributed uniformly over the entire internal surface, as shown in Figure 4. Maintaining dimensional tolerances is particularly important in solid rocket motors because local thickness variations may affect thermal protection efficiency, combustion chamber geometry, and subsequent bonding of the liner and propellant.

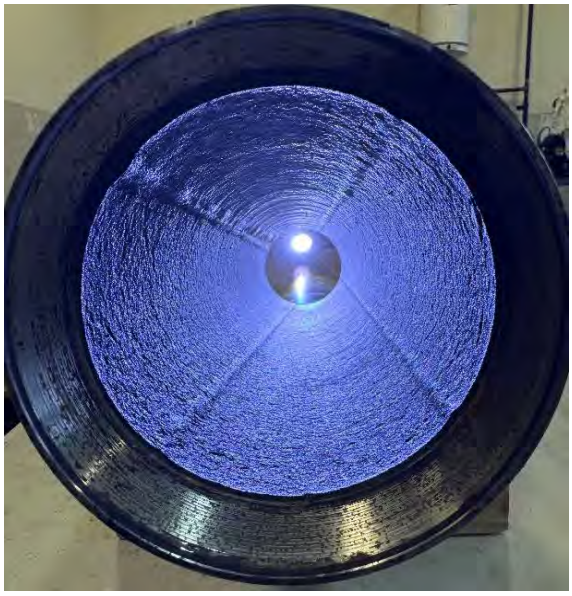


Figure 4. Interior view of the rocket motor case after application of the EPDM-based thermal insulation lining

4. CONCLUSION

An EPDM-based thermal insulation system for application to the inner wall of a steel rocket motor case was successfully developed and verified. The selected thermal insulation compound TG-H-727/3 exhibited satisfactory thermal stability, as confirmed by STA, and was therefore considered suitable for further application studies.

Adhesion testing demonstrated the effectiveness of the developed bonding system. The incorporation of the TG-HD-727/C3 adhesive layer and TG-D-727/C6 cement improved the adhesion between rubber layers and metal case. Satisfactory adhesion between the insulation system and the metal substrate was confirmed through peel and shear testing.

A technological procedure for surface preparation, application of the adhesive system, installation of the thermal insulation lining and autoclave vulcanization was successfully established. The developed procedure enabled the production of a continuous insulation lining with a thickness of 1.8–2.2 mm on the inner wall of the rocket motor case.

The successful lining and vulcanization of steel rocket motor case confirmed the applicability of the developed manufacturing technology under industrial processing conditions. The obtained results demonstrate that the proposed EPDM-based insulation system and the associated application technology are suitable for implementation in rocket motor thermal protection systems.

Acknowledgments

This work was supported by the Ministry of Science, Technological development and Innovations (Serbia), (Contract No. 451-03-34/2026-03/200325).

References

[1] Laub, Bernie, *Ablative Thermal Protection*, 2003.
[2] Bäker, M., Fiedler, T., Rösler, J., *Stress evolution in thermal barrier coatings for rocket engine applications, Mechanics of Advanced Materials and Modern Processes*, SpringerOpen, 2015.
[3] Müller, M., et al., *Thermal protection systems for liquid and solid rocket engines: materials and challenges, Aerospace*, MDP.
[4] Jiang Li, Kai Liu, Mengfei Guo, Yang Liu, Juan Wang, Xiang Lv, *Ablation and erosion characteristics of EPDM composites under SRM operating conditions*, 2018.
[5] J.W. Youren et al., *EPDM based heat shielding materials for Solid Rocket Motors: A comparative study of different fibrous reinforcements*. *Polymer Degradation and Stability*, 2013.
[6] M. Natali et al., *Effect of Wollastonite od the ablation resistance of EPDM based elastomeric heat shielding materials for solid rocket motors*, 2016.
[7] Benyahia, L., Ji, M., Poncin-Epaillard, F., *Mechanical behavior of plasma-treated metal–rubber assemblies*, *Molecules*, MDPI, 2024.
[8] J. Gržetić, J. Galović, O. Maraš, *Tehnički izveštaj VTI-04-01-1519, Osvajanje tehnologije nanošenja gumene smeše na bazi EPDM kaučuka na unutrašnji zid metalne komore raketnog motora,* 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A COMPARATIVE ASSESSMENT OF DIFFERENT NERVE AGENT VX STOCKS USING CHOLINESTERASE INHIBITION ASSAYS

SONJA BAUK
Military Technical Institute, Belgrade, bauk.sonja@gmail.com

MAJA VITOROVIĆ-TODOROVIĆ
Military Technical Institute, Belgrade, majavitod@gmail.com

MARINA ILIĆ
Military Technical Institute, Belgrade, marinailic1970@gmail.com

Abstract: *Biochemical methods for the detection of nerve agents are highly sensitive, rapid, and based on the ability of organophosphorus compounds to irreversibly inhibit enzymes from the cholinesterase family. The decrease in enzyme activity after exposure to a sample is directly proportional to the concentration of the nerve agent, making these methods powerful analytical tools, both for field detection and quantitative analysis. Nerve agents undergo various degradation pathways during long-term storage, and the complex mixture of potentially interfering compounds that accumulate over the years can affect the accuracy of enzymatic assays. In this work, a comparative analysis of the two VX samples with significantly different degradation profiles was performed using inhibition assays with electric eel acetylcholinesterase (AChE) and equine serum butyrylcholinesterase (BChE). The resulting linear regression analysis, alongside the independent GC-MS benchmark, were used to evaluate the overall inhibitory potency of the samples, screen for matrix interferences and assess the applicability of this biochemical method for quantifying residual VX in long-term stored stockpiles.*

Keywords: *cholinesterases, nerve agent VX, enzymatic method, quantitative analysis.*

1. INTRODUCTION

The interactions of highly toxic organophosphates (OPs) with enzymes from the cholinesterase family (ChEs) have been the subject of scientific research for almost a century. In this area, numerous scientific studies are focused on the development of methods for the detection and quantitative analysis of these toxic compounds. Biochemical quantification of OPs relies on the irreversible nature of ChE inhibition by OPs and is based on measuring the residual enzyme activity after a certain time of reaction with the inhibitor. Under defined experimental conditions, the reduction in enzyme activity directly correlates with both the concentration of the inhibitor and the duration of its interaction with enzyme. In certain regimes, when the inhibitor is in large excess, the inhibition reaction follows pseudo-first-order kinetics and this correlation corresponds to equation:

$$\ln (v_t/v_0) = - k_i \cdot [I] \cdot t \qquad (1)$$

where k_i is the bimolecular inhibition constant, $[I]$ is the inhibitor concentration, v_0 is the initial velocity of the enzymatic reaction and v_t is the velocity after the inhibition period t. The bimolecular inhibition constant k_i is a characteristic property of a specific irreversible inhibitor and its efficacy against a target enzyme, which serves as a key metric for comparing the inhibitory potency of different compounds [1].

Different OPs show significant differences in inhibitory potency towards ChEs from different sources. The most potent OPs known are the lethal nerve agents, which are considered to be among the deadliest chemical warfare agents to date [2]. Although their development and stockpiling as weapons are strictly prohibited by the Chemical Weapons Convention (CWC), the use of nerve agents in research, testing and development within the field of chemical defense is permitted under strict compliance with the CWC's obligations, as it remains vital for advancing protective and medical countermeasures. Consequently, possessing reliable methods for their quantitative determination is essential to maintain safety and monitor these processes effectively. While various methods for the quantification of nerve agents have been described, they often require laborious sample preparation and expensive, high-resolution analytical devices. Compared to these sophisticated techniques, biochemical methods based on ChE inhibition are rapid, simple and cost-effective [3], offering a convenient and advantageous supplementary analytical tool for many laboratory applications.

Although there are few compounds with an inhibitory potency comparable to that of nerve agents, the results of enzymatic analysis can be affected by various chemical and environmental factors. Sampling and analysis of nerve agents from different storage containers led to the identification of a large number of impurities, the type

Corresponding author: S. Bauk;
E-mail: bauk.sonja@gmail.com
DOI: 10.5937/OTEH26136B

and quantity of which depended on various factors, such as the quality and type of precursors used in production, the stabilizers applied, the manufacturing process and the numerous reaction pathways driving various degradation processes during storage [4][5]. Any potential biological activity that these compounds might exhibit towards the enzyme could affect the accuracy of enzyme inhibition-based methods. This is particularly important when the analysis is performed under conditions where the concentration of potentially interfering compounds is many times higher than the concentration of the target analyte, which is the case with highly degraded inhibitors.

Nerve agent VX (O-ethyl S-(2-diisopropylamino)ethyl methylphosphonothiolate) is one of the deadliest chemicals known. For chemical forensics, the researchers analyzed three unstabilized VX samples using NMR, LC-HRMS, and GC-MS (CI and EI) and detected over 40 compounds, including synthesis by-products, degradation products, unreacted precursors, and contaminants such as phosphonic acids, phosphonates, phosphonothioates, alkyl sulfides, aminothioalkanes, and unidentified species [4]. Consequently, enzymatic tests must take these impurities into account, especially the possibility that the samples contain a compound known as EA 2192 (diisopropylaminoethyl methylthiophosphonic acid), which is potent ChE inhibitor and almost as toxic as VX itself.

In this work, a comparative analysis of the two VX samples with significantly different purity was performed using inhibition assays with acetylcholinesterase (AChE) from electric eel and butyrylcholinesterase (BChE) from equine serum. Additionally, GC-MS analysis was performed to obtain approximate data on the quantitative ratio of the remaining VX content in the tested samples. The aim was to investigate the impact of the impurities present on the quantitative comparison of VX stocks with different degradation profile by ChE inhibition assays and to consider the application of this biochemical method for the assessment of residual VX content in long-term stored samples.

2. MATERIALS AND METHODS

Lyophilized equine serum BChE and electric eel AChE (type V-S), as well as butyrylthiocholine iodide (BTCh), acetylthiocholine iodide (ATCh) and 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB) were purchased from Sigma-Aldrich. All other reagents were of analytical grade, purchased from Sigma, Fluka or Riedel de Haën. VX samples were made available by the Military Technical Institute (Belgrade, Republic of Serbia).

Stock solutions of AChE and BChE were prepared by dissolving lyophilized enzymes in 50 mM phosphate buffer pH 7.2 (PB) and stored at 4 °C. Dilutions were made daily to obtain enzyme solutions with an approximate activity of 1 U/mL. A solution of DTNB (4.8 mM) was prepared in PB. Substrate solutions, ATCh (30 mM) and BTCh (60 mM), were prepared in distilled water. Stock solutions of VX samples (10 % v/v) were prepared in 2-propanol weekly and stored at 4 °C. On the day of the experiment, stock solutions were kept at room temperature for 1 h and than appropriately diluted in 2-propanol to obtain VX sample solutions for use in the inhibition assays.

2.1. Inhibition assays

The AChE and BChE activity measurements were performed essentially according to the method of Ellman [6], using ATCh or BTCh as the substrate, respectively, and DTNB (Ellman's reagent) as the indicator. In brief, the enzyme solution (75 µL) was added to a cuvette containing PB (2.6 mL) pre-incubated at 25 °C, followed by addition of the tested VX sample solution (25 µL). After mixing, the content was incubated for 20 min at 25 °C. The enzymatic reaction was then started by the addition of DTNB (200 µL) and corresponding substrate (100 µL). The formation of the yellow anion obtained from the reaction between Ellman's reagent and the thiocholine generated by enzymatic hydrolysis of the substrate was monitored on a Cary 60 UV/VIS spectrophotometer (Agilent Technologies) by measuring the increase in absorbance at 412 nm for 2 min at 12 s intervals. One unit of enzyme activity was defined as the amount of enzyme that catalyzes the hydrolysis of 1 μmol of substrate per minute under the described conditions.

For each enzyme, a range of dilutions of both VX samples was used to cover a broad spectrum of different inhibition levels, and results that included residual enzyme activity between 10 % and 80 % were considered for data processing. To determine the initial enzyme activity (100 %), measurements were carried out as described above, except that 2-propanol was used as the sample instead of VX solution.

Considering that the VX concentrations in the tested samples were unknown, a more convenient way to process the experimental data was to present the dependence described by equation (1) as a function of the mass concentration of the VX sample instead of the concentration of VX in the incubation mixture, which after corresponding transformation and rearrangements gives the following relationship:

$$\log(100 \cdot \frac{A_r}{A_0}) = -\frac{k_i \cdot Q \cdot t \cdot 10^{-6}}{2{,}303 \cdot R \cdot M_{VX}} \cdot C_m + 2 \qquad (2)$$

where A_r and A_0 are the residual and initial enzyme activities, k_i is the bimolecular inhibition constant ($M^{-1}min^{-1}$), Q is the purity coefficient of VX in the tested samples, t is the inhibition time (20 min), R is the dilution factor of the VX sample solutions in the incubation mixture (108), M_{VX} is the molar mass of VX (267.38 $gmol^{-1}$), and C_m is the mass concentration of the VX sample solutions (ng/mL).

Because the VX samples were a complex mixture rather than a pure compound, the obtained data points were fitted via linear regression in MS Excel using the empirical equation:

$$\log(A_r\%) = \boldsymbol{B} \cdot C_m + \boldsymbol{A} \qquad (3)$$

where A_r% is the percentage of residual enzyme activity.

The resulting regression outputs were used to evaluate and compare the overall inhibitory potency among the tested samples and to screen for potential matrix interferences caused by impurities.

2.2. GC-MS analysis

Each of the two VX samples was prepared at two nominal concentration levels, approximately 4.7 and 9.0 μg/mL, using acetonitrile as the solvent. The lower- and higher-concentration preparations were designated **S1** and **S1h** for **VX1** sample, and **S2** and **S2h** for **VX2** sample, respectively. Semiquantitative estimation of the VX content was performed using naphthalene-d8 as an internal standard at a final concentration of 2 μg/mL. The samples were analyzed using a Shimadzu GCMS-QP2020 instrument, and the chromatograms and peak heights were evaluated using LabSolutions software. To screen the chromatograms for VX and VX-related degradation products, the extracted-ion chromatogram (EIC) at m/z 114 was examined. This ion is a characteristic, high-abundance fragment of VX and related compounds containing the 2-(diisopropylamino)ethyl moiety. The instrumental conditions are summarized in **Table 1**. Splitless injection was used to improve the detection of minor sample components and to facilitate comparison of their chromatographic responses.

Table 1. GC-MS operating conditions

GC conditions	
Parameter	**Value**
Initial oven temperature	40 °C (3 min)
Oven temperature program	40-150 °C at 10 °C/min; 150-280 °C at 25 °C/min; hold at 280 °C for 10.8 min
Injector temperature	250 °C
Injection mode	Grob-type splitless injection; splitless time 0.75 min
Column	DB-5
Column dimensions	30 m × 0.25 mm i.d. × 0.25 µm
Carrier gas	Helium; linear velocity 32 cm/s
MS conditions	
Parameter	**Value**
Transfer-line temperature	280 °C
Ion-source temperature	230 °C
Detector temperature	150 °C
Electron energy	70 eV
Scan range	m/z 35-500
Scan rate	3.15 scans/s
Library	NIST 05 Mass Spectral Database

3. RESULTS AND DISCUSSION

3.1. Inhibition assays

The enzyme inhibition assays were used for comparative analysis of two VX samples from different long-term stored stocks with distinct storage-induced degradation profiles. Sample **VX1** was taken from a frequently pipetted stock stored in a glass bottle with screw cap that was opened and closed periodically, while the sample **VX2** was from the flame-sealed glass ampoule. The residual AChE and BChE activities within the range of 20–80 % after 20 min inhibition were determined using 8 different mass concentrations of each VX sample for each enzyme. The obtained data points and their linear fits according to equation (3) are shown in **Figure 1** and **Figure 2**. The resulting semi-logarithmic plots confirmed a monotonically decreasing trend in concentration-dependent residual activity for both enzymes and both tested samples. Visually, a superior inhibitory potency was observed for the **VX2** sample, while more pronounced data scattering was evident for the **VX1** sample in both enzymatic assays. This pronounced scattering is likely attributed to the presence of interfering impurities, which may dynamically interact with the enzyme or the inhibitor, destabilizing the inhibition kinetics.

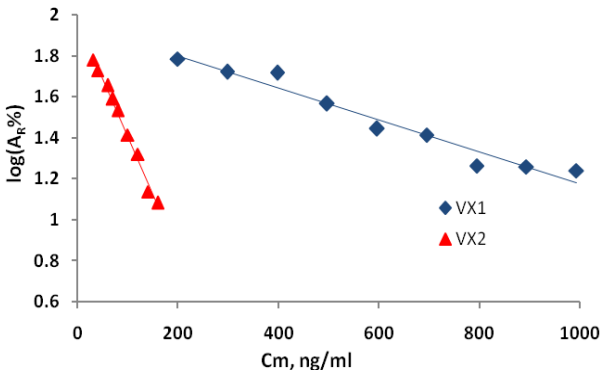


Figure 1. Semi-logarithmic plot of AChE inibition by **VX1** and **VX2** samples

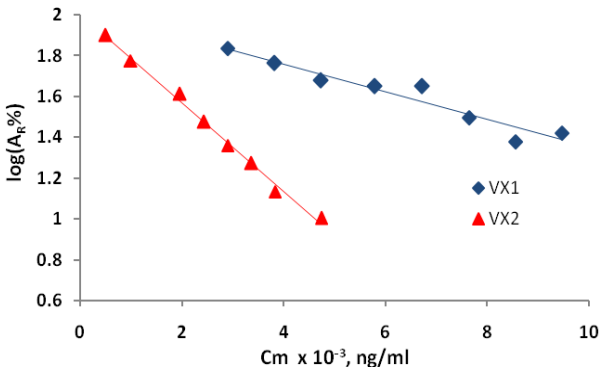


Figure 2. Semi-logarithmic plot of BChE inibition by **VX1** and **VX2** samples

The visual observations are in agreement with the expected quality of the tested samples, since the **VX1** sample originated from a source whose contents were more susceptible to handling-induced degradation and chemical alteration. In addition, AChE was more susceptible to inhibition by both tested samples, requiring much lower inhibitor concentrations to achieve the same reduction in activity as in BChE. This reveals that the AChE-based system provides a significantly higher sensitivity for the analysis of different VX samples compared to the BChE-based inhibition assay.

To further evaluate these effects quantitatively, we focused on the linear regression outputs for model presented by equation (3), which are shown in **Table 2**. The parameter ***B*** corresponds to the slope of the regression line and accounts for both the inhibitory potency and the purity factor of the tested samples, while the parameter ***A*** denotes the intercept on the ordinate and corresponds to the initial enzyme activity in the absence of the inhibitor, with theoretical value of 2. The calculated values for both parameters highly supported the visual findings and reflected the distinct inhibition profiles influenced by the purity-induced factors of the each sample. The obtained intercepts (***A***) for both enzymatic systems showed minor deviations from the theoretical value, confirming that the model closely reflects the initial enzyme activity before the inhibition. The slope values (***B***) corresponded to a higher sensitivity of AChE compared to BChE and a superior potency of **VX2** sample over **VX1**, the latter implying a higher purity of the **VX2** sample. The lower coefficients of determination (R^2) and higher standard errors (***SE***) associated with **VX1** sample successfully quantified the higher degree of experimental noise and data scattering observed visually, likely driven by the presence of interfering impurities. Conversely, the high R^2 values and lower ***SE*** values observed for **VX2** across both enzymes pointed towards less interference caused by impurities compared to **VX1**.

Table 2. The linear regression outputs: the slope (***B***), the intercept (***A***), standard errors of ***B*** and ***A*** (SE_B, SE_A) and the coefficient of determination (R^2)

ChE/sample	$B \pm SE_B$	$A \pm SE_A$	R^2
BChE/VX1	$(-6.77\pm0.77)\times10^{-5}$	2.029±0.050	0.93
BChE/VX2	$(-21.68\pm0.67)\times10^{-5}$	2.003±0.020	0.99
AChE/VX1	$(-7.76\pm0.62)\times10^{-4}$	1.953±0.040	0.96
AChE/VX2	$(-56.26\pm2.03)\times10^{-4}$	1.969±0.019	0.99

Linearization of the data for concentration-dependent inhibition was based on the model represented by equation (2), which was derived assuming pseudo-first order conditions, the presence of only one active inhibitory species (VX), and the absence of any interference caused by impurities in VX samples. If all of the above assumptions were satisfied, the slope ratio of samples **VX2** and **VX1** (B_{VX2}/B_{VX1}) would be the same for both ChE assays, as would the slope ratio determined by AChE and BChE assays (B_{AChE}/B_{BChE}) for each tested sample. This follows from the variables that define the slope in ideal conditions according to equations (2) and (3), since the expression for the slope ratio would reduce to the purity ratio (Q_{VX2}/Q_{VX1}) in the first case and to the bimolecular inhibition constant ratio ($k_{i(AChE)}/k_{i(BChE)}$) in the second.

To evaluate whether these enzyme inhibition assays can be used to quantitatively compare the VX content in tested samples, the ratio of the regression slopes B_{VX2}/B_{VX1} was calculated for each enzyme. Under ideal pseudo-first-order kinetics with a single active inhibitory component and no matrix interference, this ratio should directly reflect the relative purity of the two samples and remain constant regardless of the enzyme used. However, the experimental data (**Table 3**) contradict this ideal model, as they revealed a significant discrepancy between the values obtained by AChE and BChE assays. The facts that AChE-based test yielded a slope ratio that was more than twice the value obtained by BChE (7.25 vs. 3.20) and that the slope ratios were enzyme-dependent, strongly indicate that the ideal kinetic assumptions are not met. This discrepancy provides definitive evidence that the VX sample impurities do not behave as inert diluents, but actively engage in enzyme-specific interactions or interfere with the primary inhibitors bioavailability. These findings demonstrate that ChE inhibition assays cannot be universally and reliably applied for the direct quantitative cross-comparison of VX content across the samples of different and unknown purity, as the unpurified samples matrices actively interferes with the inhibition kinetics in an enzyme-specific manner.

Table 3. Comparative presentation of regression-derived slope ratios and the resulting calculated values: relative purity of **VX1** sample versus **VX2** sample (Q_{VX1}), and the relative k_i of BChE versus AChE ($k_{i(BChE)}$).

Enzyme	B_{VX2}/B_{VX1}	Q_{VX1} (%)
AChE	7.25	13.79
BChE	3.20	31.21
Sample	B_{AChE}/B_{BChE}	$k_{i(BChE)}$ (%)
VX1	11.46	8.72
VX2	25.95	3.85

Since the bimolecular inhibition constant is a characteristic property of each individual enzyme-inhibitor system, a biochemical fingerprint for the identification of an unknown irreversible inhibitor can be obtained by assessing its inhibitory potency against a group of structurally different or related enzymes. The inhibition spectrum (or relative inhibitory response profile), obtained by normalizing all k_i values for the enzymes in the group as a percentage of the highest value within that set, serves as a unique multi-enzyme barcode for identifying inhibitors.

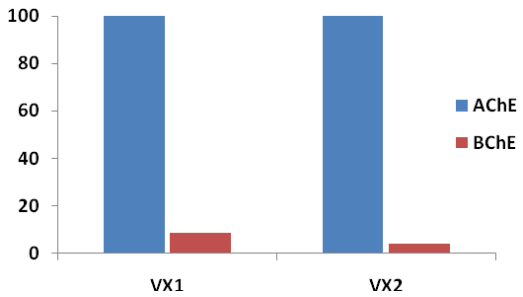


Figure 3. Inhibition spectrum of **VX1** and **VX2** samples

The two-enzyme inhibitory spectrum was established for each VX sample using relative k_i values derived from the ratio of the regression slopes between AChE and BChE. Although this approach mathematically does not provide new independent information, the graphical comparison

of these spectra **(Figure 3)** allows for visual observation of asymmetry between the two tested samples.

3.2. GC-MS analysis

The chromatogram of the **S1h** sample is shown in **Figure 4**. The total ion chromatogram (TIC) is shown in black, whereas the extracted-ion chromatogram at m/z 114 is shown in pink. Only a very small VX peak was observed at a retention time of approximately 17.350 min. Two additional relevant compounds were tentatively identified by comparison with the NIST library: *bis*[2-(diisopropylamino)ethyl] sulfide at 18.832 min, with an 81% library match, and *bis*[2-(diisopropylamino)ethyl] disulfide (DES) at 21.175 min, with an 88 % library match. The markedly lower VX response relative to the DES response qualitatively indicates extensive degradation of the **S1h** sample, only a small residual amount of the parent compound remaining.

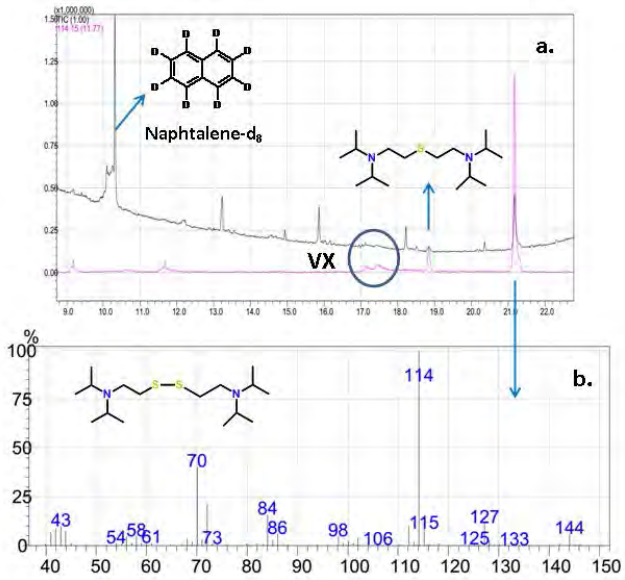


Figure 4. (a) Chromatogram of the **S1h** sample: TIC (black) and EIC at m/z 114 (pink); **(b)** mass spectrum of *bis*[2-(diisopropylamino)ethyl] disulfide (DES).

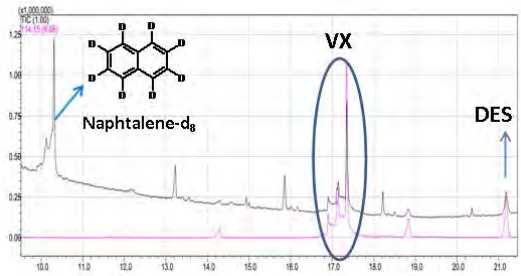


Figure 5. Chromatogram of the **S2h** sample: TIC (black) and EIC at m/z 114 (pink).

In the **S2h** chromatogram (**Figure 5**), VX was the dominant component, accompanied by smaller peaks tentatively assigned to DES and the corresponding sulfide. Pronounced tailing of the VX peak was observed, likely reflecting non-optimized chromatographic conditions. To assess whether the relative VX contents could be compared semi-quantitatively, the VX-to-internal-standard peak-height ratios were calculated for the **S2** and **S2h** samples. The ratio between these responses was consistent with the ratio of the nominal sample concentrations, supporting the use of this approach for a rough relative comparison under the applied conditions. On this basis, the residual VX contents in **S1** and **S1h** samples were estimated to be approximately one-fifth of those in the corresponding **S2** and **S2h** samples, indicating a relative purity of 20 % for the **VX1** sample compared to the **VX2** sample (Q_{VX1}). Several additional peaks observed in both chromatograms at retention times of approximately 13.2, 15.9, and 18.2 min did not contain the m/z 114 ion and were tentatively attributed to long-chain hydrocarbon contaminants originating from the GC column or inlet septum.

3.3. The practical relevance of ChE inhibition assays and future perspectives

The quantitative ratio of the remaining VX content in the tested samples determined by the slope-ratio method using AChE, BChE and by GC-MS analysis was found to be 7.25, 3.20 and 5, respectively, corresponding to a relative purity of the **VX1** sample of 14 %, 31 % and 20 %.

Despite the observed discrepancies between enzymatic assays and reference GC-MS analysis, the obtained results have significant practical and predictive value for analytical applications. This significance is manifested trough two key aspects, the first of which relates to the validity of the method as a screening tool, and the second to the dependence of the assay accuracy on the VX content ratio in the samples and their absolute purity.

Regarding the former, both applied enzymatic assays consistently identified the **VX1** sample as unusable for R&D applications. The results show that VX content in the **VX1** sample is below 31 % relative to the sample of **VX2**, so its absolute purity cannot exceed 31 %, the maximum value that can be reached if the **VX2** sample were 100 % pure. From an analytical perspective, whenever the determined quantitative ratio between samples exceeds 3, unambiguous information regarding the unsuitability of at least one sample is generated. In this context, the numerical discrepancies between the methods lose relevance, as the primary objective of rapid screening of inappropriate samples is successfully achieved. Therefore, these enzymatic assays provide a clear, rapid diagnostic signal to disqualify low-purity samples without immediately relying on cost-intensive chromatographic techniques.

Considering the second aspect, it is reasonable to expect that the convergence of results among different methods will increase as the difference in the quantitative inhibitor ratio decreases. However, this trend is a direct function of the absolute purity of the tested samples, which dictates the profile and variance of the interfering matrix impurities. Assuming that the ratio of residual VX content in the tested samples was exactly 5, the feasibility of using these enzyme assays to obtain analytically acceptable results can be evaluated through two extreme hypothetical scenarios, which are presented below.

In the high-purity scenario (e.g., samples of **VX2** and **VX1** had purity of 95 % and 19 %), the difference in the impurity matrix was extreme (5 % vs. 81 %), explaining the pronounced deviations in the enzymatic assays. In this case, samples with a lower quantitative ratio (< 5), regardless of their absolute purity, may only have a reduced difference in impurity levels. Therefore, it can be predicted that their testing would inherently lead to analytically acceptable agreement between the two enzyme assays, at the point when their quantitative ratio falls below a certain limit.

In the low-purity scenario (e.g., samples of **VX2** and **VX1** had purity of 5 % and 1 %), the impurity fractions were inherently similar (95 % vs. 99 %). Reducing the quantitative ratio would reduce the difference in impurity levels, but only if the absolute purity of the tested samples remain at the level of the low purity scenario. However, samples with higher absolute purity and the same quantitative ratio of 5 may only have an increased difference in impurity levels and consequently a greater difference in enzyme assay results. In this case, the performance of assays at ratios below 5 for a highly purified sample remains an open question.

Given that GC-MS analysis confirmed a significantly higher abundance of degradation products in **VX1** sample, the low-purity scenario can be rejected. This finding, supported by the results of ChE inhibition tests which pointed at high impurity loads in **VX1** sample, implies that the test conditions were closer to the high-purity scenario. Accordingly, the assumption that lower quantitative ratios of the tested VX samples will yield improved correlation in the enzymatic assays is well-founded. To fully define the accuracy windows and practical boundaries of enzymatic inhibition assays in VX sample analysis, further systematic investigations and development of matrix-matching calibration strategy are required. The focus of future work should be on establishing a robust calibration framework and selecting the most resilient and suitable single enzyme for high-throughput, rough quantitative profiling of partly-degraded VX samples.

5. CONCLUSION

The aim of this work was to examine the influence of impurities, present in two VX samples with different degradation profiles, on the results of ChE inhibition assays and their application in the quantitative comparison of residual VX content in stocks that have undergone distinct storage-induced decomposition.

The experimental results strongly supported the initial assumption regarding the varying degrees of purity between the two tested samples. The linear regression analysis for the inhibition of AChE and BChE yielded distinct profiles regarding both inhibitory potency and data consistency, driven by purity-related matrix interferences in each sample. The calculated relative purity values for **VX1** sample determined by AChE and BChE inhibition assays correctly classified the sample as heavily contaminated and unsuitable for application, a finding fully supported by preliminary GC-MS data. The dual-enzyme screening approach demonstrated excellent feasibility for ensuring the accuracy and reliability of enzyme-based assays during the quantitative determination of target inhibitors in non-homogeneous samples. The combination of regression metrics and slope ratio divergence across two distinct enzymes served as an effective tool to detect the presence of interfering impurities, assess relative sample heterogeneity and evaluate the susceptibility of different target enzymes to matrix effects.

The proposed methodology provides a robust and effective analytical framework for sample triage, preliminary quality assessment and semi-quantitative estimation of relative inhibitor content prior to definitive chromatographic analysis. While GC-MS remains indispensable for the final, precise quantification of highly heterogeneous matrices, the enzymatic approach serves as an excellent first-line screening method. Definitive mapping of the applicability threshold, which includes working with a broader sample library with varying purity gradients, a wider panel of alternative enzymes, and systematic cross-referencing with GC-MS data, will optimize this inhibition assay into a standardized screening protocol, fully capable of achieving a high degree of accuracy within analytically acceptable results.

References

[1] ROBERT A. COPELAND, R.A.: *Evaluation of Enzyme Inhibitors in Drug Discovery: a guide for medicinal chemists and pharmacologists*, John Wiley & Sons, Hoboken, New Jersey, 2013.

[2] POHANKA, M., ADAM, V., KIZEK, R.: *An Acetylcholinesterase-Based Chronoamperometric Biosensor for Fast and Reliable Assay of Nerve Agents*, Sensors, 13 (2013) 11498-11506.

[3] SCHWAB, A., HORN, G., KEHE, K., WOREK, F., AMEND, N.: *A simple acetylcholinesterase inhibition assay for the quantifcation of the nerve agent VX: Application in a Franz cell model with rat skin and various decontaminants*, Toxicology Letters, New York, 404 (2025) 20–27.

[4] OVENDEN, S.P.B., WEBSTER, R.L., MICICH, E., McDOWALL, L.J., McGILL, N.W., WILLIAMS, J., ZANATTA, S.D.: *The identification of chemical attribution signature of stored VX nerve agent using NMR, GC-MS, and LC-HRMS*, Talanta 211 (2020) 120753.

[5] WEBSTER, R.L., OVENDEN, S.P.B., McDOWALL, L.J., DENNISON, G.H., LAWS, M.J., McGILL, N.W., WILLIAMS, J., ZANATTA, S.D.: *Chemical forensic profiling and attribution signature determination of sarin nerve agent using GC-MS, LC-MS and NMR*, Analytical and Bioanalytical Chemistry, 414 (2022) 3863-3873.

[6] ELLMAN, G.L., COURTNEY, K.D., ANDRES, JR.V., FEATHERSTONE, R.M.: *A new and rapid colorimetric determinaton of acetylcholinesterase activity*, Biochemical Pharmacology, 7(2) (1961) 88–95.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MAGNETOIMPEDANCE EFFECT OF $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ AMORPHOUS RIBBON

RADOSLAV SURLA
Military Technical Institute, Belgrade, Serbia, ekorade@gmail.com

JOVANA MANDIĆ
Military Technical Institute, Belgrade, Serbia, jmandic96@gmail.com

JELENA ORELJ
Faculty of Technical Sciences, University of Kragujevac, Serbia, jelena.orelj@ftn.kg.ac.rs

NEBOJŠA MITROVIĆ
Faculty of Technical Sciences, University of Kragujevac, Serbia, nebojsa.mitrovic@ftn.kg.ac.rs

PAVEL CRNOMARKOVIĆ
Military Technical Institute, Belgrade, crnomarkovic.pavel@gmail.com

Abstract: *The paper presents the results of an experimental investigation of the magnetoimpedance (MI) response of an $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ amorphous ribbon, prepared by the melt-spinning rapid quenching technique on a rotating copper wheel. Characterization revealed that the internal configuration and the profile of the MI curves exhibit a strong frequency dependence, reaching a maximum MI ratio value of 31.3% at an operating frequency of 36.11 MHz without an external magnetic field. With an increase in the magnetic field intensity, a continuous decrease in the MI ratio values is observed. At frequencies above 75 MHz, slight distortions in the profile of the experimental curves are noticed at a field strength of approximately 8 kA/m, which hints at the presence of an internal magnetic anisotropy field. Considering the completely amorphous structure of the sample in the as-cast state, this aspect cannot be analyzed in more detail at this stage; however, its clear expression and dominant influence on the soft magnetic properties are expected after the thermo-activated transformation of the amorphous matrix into a nanocrystalline state via controlled isothermal annealing.*

Keywords: *Magnetoimpedance effect, Amorphous alloys, Melt-spinning, Magnetic anisotropy.*

1. INTRODUCTION

The phenomenon of magnetoimpedance (MI) represents a specific electromagnetic response observed in a narrow group of soft magnetic alloys with an amorphous or nanocrystalline structure. This effect is manifested through an intensive variation in the complex impedance of a ferromagnetic sample under the influence of an external direct current (DC) magnetic field. Since its discovery at the end of the last century, this phenomenon has become a fundamental platform for the development of high-sensitivity magnetic field sensors used in automotive, military, and aerospace industries [1-8].

It has been experimentally proven that the impedance module of a ferromagnetic material decreases with an increase in the external field intensity, while simultaneously increasing with the frequency of the excitation current flowing through the sample. Quantification of this effect is most commonly performed by defining the MI ratio according to the following equation:

$$\Delta Z/Z(\%)=100\% \cdot (Z(H)-Z(H_{max}))/Z(H_{max}), \quad (1)$$

where $Z(H_{max})$ represents the reference impedance value measured at the maximum intensity of the applied external magnetic field. Depending on the internal magnetic configuration and processing technology, under zero-field conditions (H = 0), the influence of remanent magnetization may come into play, the disappearance of which with an increasing field causes a characteristic asymmetry in the MI diagrams for opposite directions of the applied field (for the -H and +H directions).

The frequency dependence of the magnetoimpedance is directly induced by the skin effect, which is described by the equivalent penetration depth of the current through the conductor. This value increases due to the drop in the total magnetic permeability in a strong magnetic field, while it decreases with increasing frequency. The action of a high-frequency alternating current induces strong eddy currents in the sample, causing the displacement of charge carriers and the concentration of electrical energy towards the peripheral regions of the conductor. The penetration depth is defined by the following relation:

$$\delta_m=(\rho/(\pi\cdot\mu\cdot f))^{0.5}, \quad (2)$$

where σ is the electrical conductivity, μ is the magnetic permeability, and f is the frequency of the alternating current. In addition to pure electromagnetic parameters, the MI response is crucially influenced by the domain structure of the magnetic material, which is a direct function of the internal structural arrangement and the applied sample geometry. Therefore, the maximum values of the MI ratio

Corresponding author: R.Surla;
E-mail: ekorade@gmail.com
DOI: 10.5937/OTEH26137S

represent a compromise between the frequency and the strength of the magnetic field, which is precisely defined by experimental measurements.

2. EXPERIMENTAL PROCEDURE

The subject of experimental investigations in this study is a multi-component alloy with the composition $Fe_{72}Al_5Ga_2P_{11}C_6B_4$. Samples were synthesized in the form of thin ribbons (geometric dimensions: width about 2 mm, thickness about 40 μm) by applying the technology of ultra-rapid quenching of the liquid alloy melt on a rotating copper wheel (melt-spinning method) in a protected atmosphere.

The characterization of the magnetoimpedance (MI) effect was performed in a homogeneous external magnetic field with a maximum intensity up to 15 kA/m, which was generated using a custom-designed and controlled system with two pairs of Helmholtz coils providing field homogeneity above 99% [5, 9]. The measurement of the complex impedance of the investigated ribbon samples was realized in a wide frequency range using a vector network analyzer (instrument: Vector Network Analyzer Agilent 8753ES).

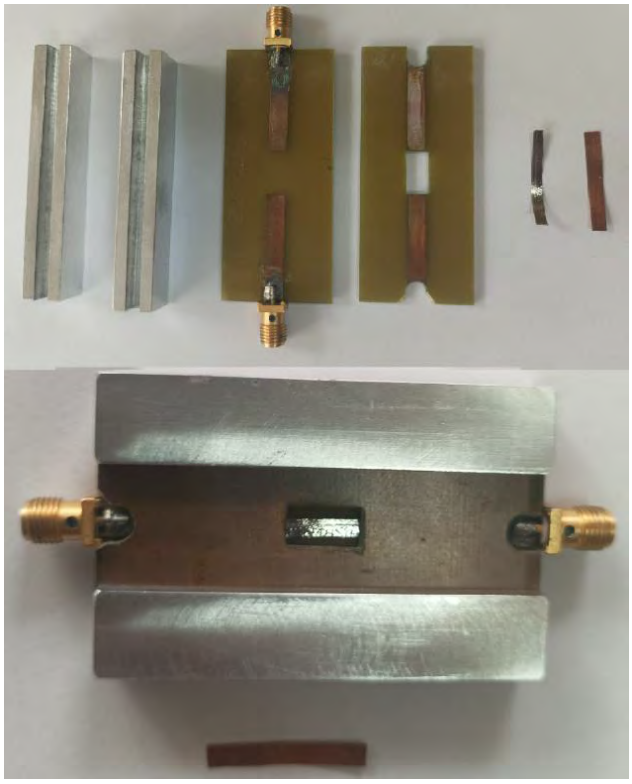


Figure 1. Sample holder configuration: (top) components of the microstrip holder, (bottom) amorphous ribbon sample positioned within the holder, showing the short-circuit element used for network analyzer calibration.

The sample holder was specifically designed in the form of a microstrip to accommodate the specific geometry of the investigated ribbon (Figure 1). The ends of the electrical contacts were placed at a fixed distance of 1 cm, while the total length of the measured ribbon was at least 2 cm to eliminate the influence of contact resistances at the connection points. The measurement is realized by connecting to a single port of the instrument, which automatically converts the reflection S_{11} parameters into the impedance of a 1 cm long sample. The systematic error introduced by the cables and the holder itself was eliminated by precise calibration of the measurement setup using a copper short-circuit jumper of similar dimensions.

The structural properties of the obtained ribbons in the initial state were investigated by X-ray diffraction (XRD) on a Rigaku SmartLab apparatus using a Cu Kα radiation source. The identification of the amorphous nature of the material was performed based on the shape and geometry of the diffraction lines recorded at room temperature. The surface morphology and topography of the free side of the amorphous ribbon were analyzed using scanning electron microscopy (SEM JEOL JSM-6610-LV), while the chemical homogeneity and quantitative composition of the constituent elements of the alloy were verified by energy-dispersive X-ray spectroscopy (EDS).

3. RESULTS AND DISCUSSION

3.1 XRD Analysis

X-ray structural analysis (XRD) confirmed the amorphous character of the synthesized ribbon in the initial as-cast state. In the obtained diffractogram (Figure 2), only a characteristic broad diffraction maximum (the so-called amorphous halo) is observed in the 2θ angle range between 40° and 50°. The complete absence of any sharp and intense Bragg reflections clearly indicates the absence of long-range crystalline order within the investigated material. This appearance of the diffraction profile unequivocally confirms that the applied ultra-rapid melt quenching technology successfully suppressed premature crystallization processes and effectively froze the structural disorder in the sample.

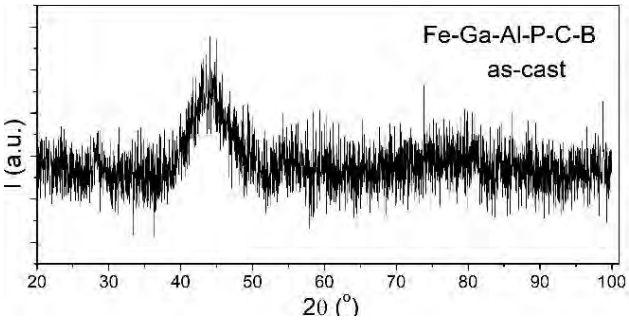


Figure 2. XRD pattern of the $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ amorphous ribbon in the as-cast state.

3.2 SEM and EDS Analysis

Scanning electron microscopy (SEM) was applied for direct investigation of the surface morphology of the free (shiny) side of the cast ribbon. The micrograph shown in Figure 3 demonstrates high structural homogeneity and surface smoothness, without the presence of microporosity, cracks, or early crystal nuclei that could degrade the magnetic softness of the ribbons.

Concurrently, EDS analysis (Figure 4) was used for detailed chemical characterization of the sample. The results of quantification in atomic and weight fractions,

summarized in Table 1, confirm that the experimentally determined composition completely corresponds to the nominal stoichiometry of the $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ alloy, testifying to the high chemical uniformity of the material along the entire ribbon sample.

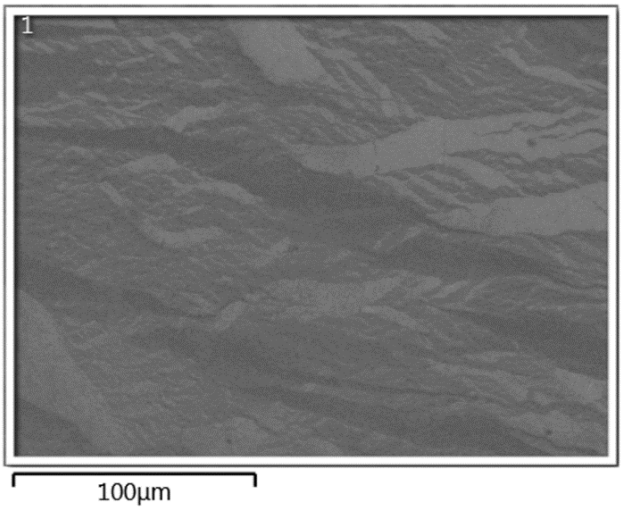


Figure 3. SEM micrograph of the free (shiny) side of the $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ amorphous ribbon.

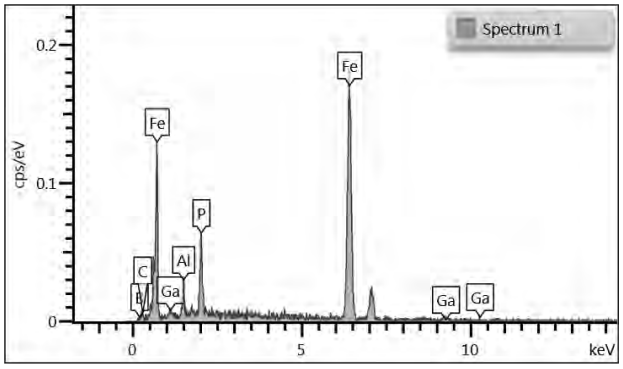


Figure 4. SEM-EDS spectrum and elemental analysis of the as-cast $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ amorphous ribbon.

Table 1. Quantitative elemental composition (weight and atomic percentages) of the investigated $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ amorphous ribbon.

Element	Wt%	Atomic %
B	3.00	11.68
C	3.25	11.37
Al	3.49	5.44
P	7.11	9.64
Fe	78.43	59.03
Ga	4.71	2.84
Total:	100.00	100.00

3.3 Magnetoimpedance Measurements

The diagram in Figure 5 shows the MI ratio as a function of the excitation current frequency for different values of the external magnetic field. Experimental characterization established that the MI effect is highly pronounced in the frequency interval from 17 MHz to 40 MHz. The MI ratio reaches its maximum value at a frequency of 36.11 MHz under zero external field conditions (H = 0 A/m) and amounts to 31.3%.

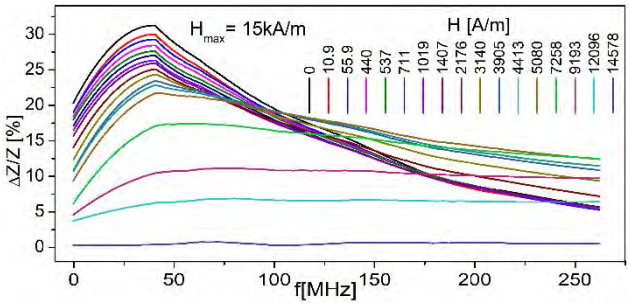


Figure 5. Frequency dependence of the MI ratio for different external magnetic field strengths up to H_{max} = 15 kA/m.

Analysis of the diagrams in Figures 5 to 8 reveals a constant, monotonic decreasing trend of the MI ratio with an increase in the magnetic field intensity acting longitudinally on the measured sample. This decrease in impedance is a consequence of the rotation of the magnetization vector in the direction of the external magnetic field. Also, the profile of the MI curves shows constant growth with increasing frequency up to 36.11 MHz (Figure 6), while with a further increase in frequency above this value, the curves are in a noticeable decline (Figure 8).

At frequencies higher than 75 MHz, specific deformations and distortions of the experimental curve profiles are observed at an external magnetic field strength of around 8 kA/m. Considering that the investigated alloy is completely amorphous in the initial state, this phenomenon cannot be analyzed or quantified in detail at this stage of the research; instead, the mentioned distortions serve as an early indication that only hints at the presence of an internal magnetic anisotropy field. This anisotropy is currently overshadowed by internal stresses generated during the rapid cooling of the liquid alloy melt [10 - 12].

Further insight into the frequency range from 18 to 36 MHz shows that the curves are almost completely overlapped (Figure 7). Therefore, it is expected that a controlled thermo-activated treatment (annealing) will lead to structural relaxation and transformation of the amorphous matrix into a two-phase nanocrystalline structure (with a dominant α-Fe(Si) phase), which will enable a clear expression of the magnetic anisotropy field and a drastic improvement in the MI performance, which will be the subject of our future work.

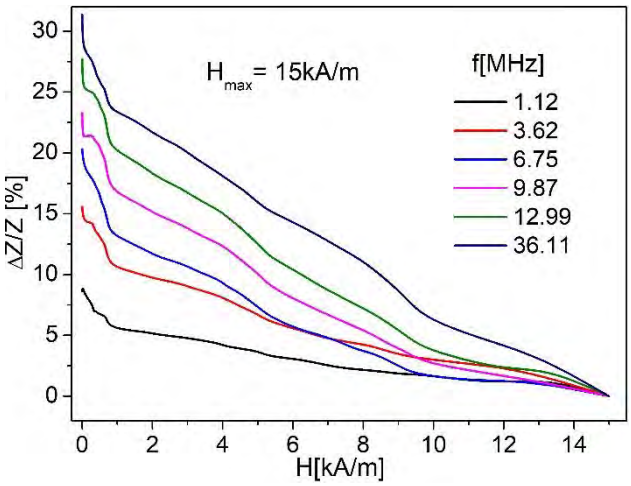


Figure 6. Magnetic field dependence of the MI ratio profiles at low excitation current frequencies ranging from 1.12 MHz to 36.11 MHz.

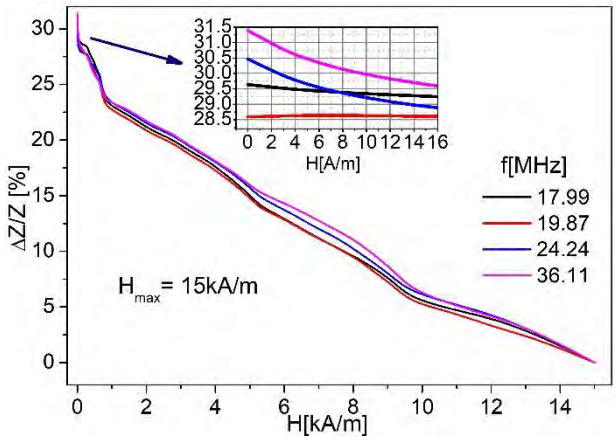


Figure 7. Superposition and convergence of the MI ratio curves in the optimal frequency range from 17.99 MHz to 36.11 MHz.

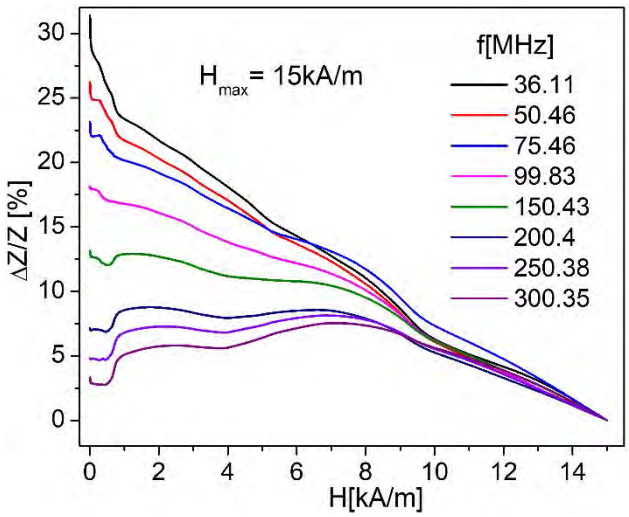
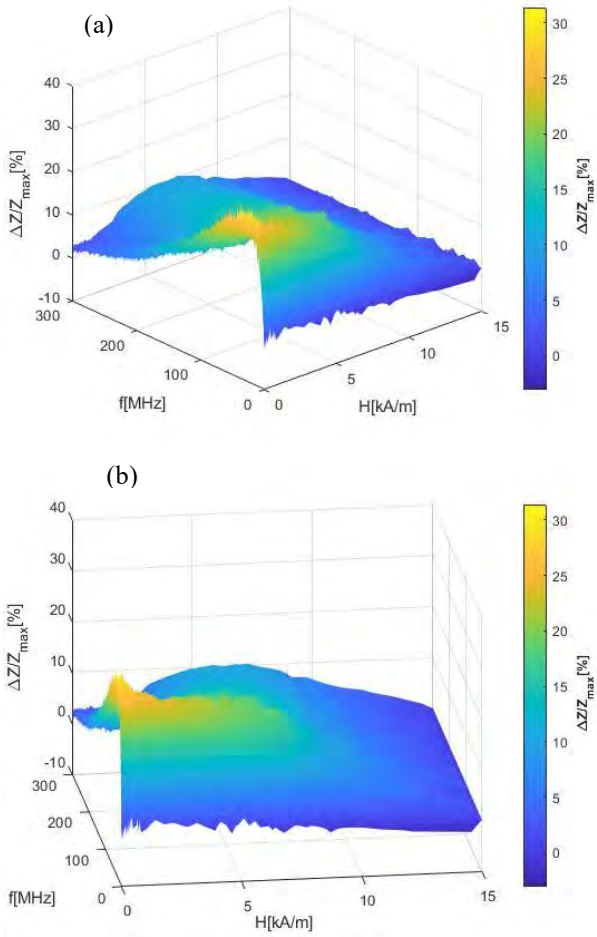


Figure 8. Magnetic field dependence of the MI ratio profiles at high frequencies from 36.11 MHz to 300.35 MHz, showing profile distortions near 8 kA/m that hint at the internal magnetic anisotropy.

In order to comprehensively evaluate the simultaneous influence of the excitation current frequency and the external magnetic field strength on the magnetoimpedance properties of the investigated alloy, three-dimensional dependence diagrams along with a corresponding two-dimensional projection map were constructed, as presented in Figure 9. On the three-dimensional surface plots (Figures 9a and 9b), the spatial profile of the MI response is clearly and explicitly observed across the entire measured frequency range up to 300 MHz. This representation allows an unequivocal insight into the curve evolution, where a dominant peak of the MI ratio stands out at 36.11 MHz under low external field conditions, followed by a symmetrical drop in impedance with increasing field strength due to the magnetic saturation of the sample. On the other hand, the 2D projection color map from the top view (Figure 9c) provides a precise topographic overview of the impedance contours, where the optimal frequency zone for the stabilization of the effect is distinctly demarcated colorimetrically. These visualizations confirm the high linearity and sensitivity of the MI response in the characteristic domain, establishing a crucial experimental foundation for the precise engineering of future magnetic sensors based on this amorphous ribbon.



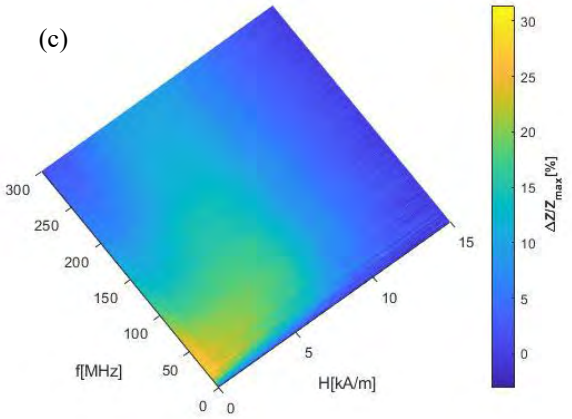


Figure 9. Magnetoimpedance (MI) ratio profiles as a multi-variable function of the excitation current frequency ,f and external magnetic field strength**: (a)** 3D surface plot from the front view emphasizing low-frequency peaks, **(b)** 3D surface plot from the rear view illustrating the high-frequency drop and profile distortions, and **(c)** 2D projection color map (top view) highlighting the optimal operational frequency window and convergence zones.

5. CONCLUSION

In this work, the magnetoimpedance response of an amorphous ribbon of a multi-component $Fe_{72}Al_5Ga_2P_{11}C_6B_4$ alloy obtained by the melt-spinning method was successfully evaluated. It was experimentally confirmed that the investigated material achieves its optimal MI profile at a characteristic frequency of 36.11 MHz, where the value of the MI ratio is 31.3% in the absence of a field. The shape and trend of the obtained curves show a high sensitivity to changes in the frequency of the excitation signal.

At high frequencies above 75 MHz and at a field strength of 8 kA/m, specific curvatures of the experimental curves were observed, representing an early indication and a hint of the existence of a magnetic anisotropy field. Since this phenomenon is overshadowed by structural disorder and residual casting stresses in the initial state, a complete analytical picture of the anisotropy field, as well as the projected increase in magnetic softness and sensitivity of the sensor element, are planned in the next phase through controlled isothermal annealing and the generation of a stable nanocrystalline matrix.

Acknowledgement

This work was supported by the Ministry of science, technological development and innovations (Serbia), Contract No. 451-03-34/2026-03/200325.

References

[1] Panina, L.V., Mohri, K.: *Magneto-impedance Effect in Amorphous Wires*, Applied Physics Letters, 65(9), (1994), 1189 – 1191.

[2] Kraus, L.: *Theory of Giant Magneto-impedance in the Planar Conductor with Uniaxial Magnetic Anisotropy*, Journal of Magnetism and Magnetic Materials, 195(3), (1999), 764 – 778.

[3] Surla, R., Mitrović, N., Đukić, S., Ibrahimović, V.: *Amorphous $Fe_{72}Cu_1V_4Si_{15}B_8$ Ribbon as Magneto-Impedance Sensing Element*, Serbian Journal of Electrical Engineering, 13(3), (2016), 381-394.

[4] Surla, R., M., Vasić, Mitrović, N., Radović, Lj., Minić, D.: *Magnetic properties of* $Fe_{72}Cu_1V_4Si_{15}B_8$ *alloy with a composite amorphous/ nanocrystalline structure,* Journal of Magnetism and Magnetic Materials, 564 (2022) 170141

[5] SURLA, R.: Doktorska disertacija: *Uticaj odgrevanja na strukturne transformacije i magnetna svojstva legure $Fe_{72}Cu_1V_4Si_{15}B_8$,* FTN Čačak, Univerzitet u Kragujevcu, 2021.

[6] Vasić M., Surla R., Mitrović N., Minić Dragica., Mitrović N., Maričić A., Minić Dušan: *Thermally induced microstructural transformations of $Fe_{72}Cu_1V_4Si_{15}B_8$ alloy*, Metal. Mater. Trans. A 48 (2017) 4393–4402.

[7] Surla, R., M., Vasić, Mitrović, N., Minić, D.: *Synthesis and characterization of $Fe_{72}Cu_1V_4Si_{15}B_8$ metallic glass ribbons prepared by melt spinning casting*, 9th International Scientific Conference on Defensive Technologies, Belgrade, 8-9, October 2020.

[8] Surla, R., Mitrović, N., Orelj, J., Crnomarković, P.: *Magnetoimpedance effect of nanocrystalline FeNiSiB and FeCuVSiB ribbons*, 11th International Scientific Conference on Defensive Technologies, Belgrade, 9-11, October 2024.

[9] Surla, R., Mitrović, N., Crnomarković, P.: *Designing the controlled generator of a homogeneous magnetic field*, *10th International Scientific Conference on Defensive Technologies*, Belgrade, 13-14, October 2022, ID 116, ISBN 978-86-81123-85-0.

[10] Herzer, G.: *Grain structure and magnetism of nanocrystalline ferromagnets*, IEEE Transactions on Magnetics, 25 (1989) 3327-3329.

[11] Kulik, T.: *Nanocrystallization of metallic glasses*, Journal of Non-crystalline Solids, 287 (2001) 145-161.

[12] Herzer, G.: *Anisotropies in soft magnetic nanocrystalline alloys*, Journal of Magnetism and Magnetic Materials, 294 (2005) 99-106.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CORRELATION BETWEEN IMPACT ENERGY AND THICKNESS IN QUENCHED AND TEMPERED LOW ALLOYED STEEL

MILOŠ ČENIĆ
National Institute of Chemistry, Ljubljana, Slovenia, Milos.Cenic@ki.si

ESMAIL ALI SALEM AHMED

College of Civil Aviation and Meteorology, Tripoli, Libya,
University of Belgrade, Faculty of Technology and Metallurgy, Serbia, Belgrade esmail_682005@yahoo.com

SRDJA PERKOVIĆ
Military Technical Institute, Belgrade-Zarkovo, srdja.perkovic@mod.gov.rs

NENAD RADOVIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, nenrad@tmf.bg.ac.rs

DRAGOMIR GLIŠIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, gile@tmf.bg.ac.rs

STEFAN DIKIĆ
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, sdikic@tmf.bg.ac.rs

Abstract: *The aim of this work was to establish the correlation between thickness and impact energy. The steel tested in this work was 42CrMo4 steel in quenched and quench and tempered condition. Impact energy was measured on impact Charpy tester for specimens with width of 10 mm and thickness of 5,0mm, 7,5 mm and 10,0 mm. After austenitization of 850˚C/15minutes and quenching in oil, one group of samples were tested, while other after subsequent tempering of 560˚C/90minutes. For quenched samples are characterized with brittle fracture and impact toughness of 6,8 J, 9,1 J and 14.2 J for 5,0mm, 7,5 mm and 10,0 mm samples respectively. For tempered samples are characterized with mixed ductile and brittle fracture and impact toughness of 34,4 J, 56,2 J and 69,1 J for 5,0mm, 7,5 mm and 10,0 mm samples respectively. In both groups, cross section of samples was increased linearly, but impact energies did not follow the same trend. This behavior was attributed to permanent presence of plastic deformation, no matter the microstructure, i.e. absence of proportionality between cross-section surface and plastically deformed areas. Also, this feature is more pronounced in quenched samples, due to very low toughness of martensite.*

Keywords: *Charpy notch toughness, thickness, fracture.*

1. INTRODUCTION

Steel is the most dominant structural material with annual production of more than 1.85 milions of tonnes [1]. At the same time, continuos development of new grades of steels takes place. Improvement of strength and toughness is the only approach to decrease cross sections of steel structures, enhancing decrease in weight of construction itself [2]. These trends lead to design of steel sections, pipes etc. with smaller thickness. Today, structural steels are produced in hot rolling with thickness of as low as 1.5 mm [2-4].

Strength of new steels is very well controlled by strengthening mechanisms, through roles of microalloying elements and hot rolling parameters [5]. On the other hand, toughness is parameter related to energy release and is much more statistically described. During Charpy notch impact testing, samples are designed to localize crack initiation on the tip of machined groove [5-8]. Therefore, start of crack initiation depends both on quality of the notch (machining) and microstructure and impurities in steel at the tip. This is the reason why all these details are clearly described in Standards [9].

The aim of this paper is to test the same steel in two conditions with different sample thickness and compare obtained impact energies.

2. EXPERIMENTAL PART

Chemical composition of steel 42CrMo4 tested in this work is given in table 1. Steel was forged, and machined. Two groups of samples were tested, one after quenching (austenitization at 850˚C/15minutes and quenching in oil) and another after quenching and tempering (austenitization at 850˚C/15minutes followed by oil quenching and tempering at 560˚C/90minutes).

Corresponding author: N.Radović;
E-mail: nenrad@tmf.bg.ac.rs
DOI: 10.5937/OTEH26138C

Table 1. Chemical composition of tested steel [10]

	C	Mn	Cr	Mo	P	S	Si
min	0,39	0,65	0,75	0,15			0,15
max	0,46	1,10	1,20	0,25	0,035	0,040	0,30

Charpy impact testing. The testing was performed using the SCHENCK TREBELL 150/300 J instrumented Charpy pendulum (Instron GmbH, Germany).

In order to evaluate the influence of different thickness on Charpy impact toughness samples had standardized [9] thickness of t=5,0mm, t=7,5mm and t=10,0mm, according to sketch shown in Figure 1.

Observation of fractured surface was done using stereo microscope.

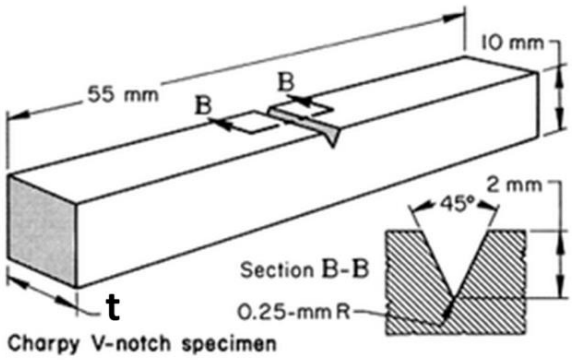


Figure 1. Sketch of Charpy impact specimen [7-8].

3.RESULTS

Fractured surface. Typical fractured surfaces obtained in quenched samples nad tempered samples are shown in Figure 2 and 3 respectively. Measured Charpy impact toughness for all specimens are given in Tables 2 and 3, respectively.

Table 2. Charpy impact energies for quenched samples

Thickness t, mm	Quenched, CVN, J	
	measured	Area normalized
5,0	6,8	6,8
7,5	9,1	10,2
10,0	14,2	13,6

In all three cases, fractured surface of quenched samples are flat and perpendicular to cross section. Crack initiation places are designated by arrow in each fractograph. Starting from crack initiation places, a network of river patterns stretch out (extend) to the edges of each sample [11,12].

At the lower edges of the samples, a small surface indicating presence of shear lips is present, and designated with white arrow on Figure 2.

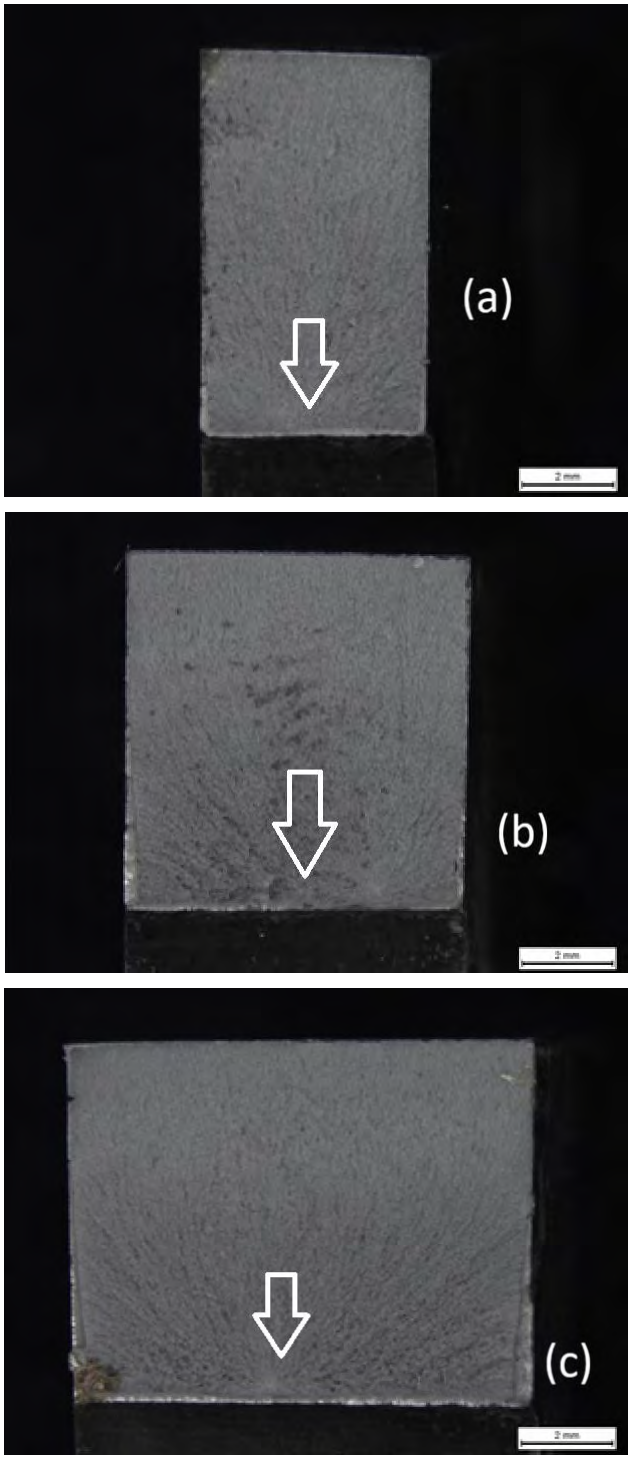


Figure 2. Fractographs of fractured surface of quenched Charpy impact specimens: (a) thickness 5,0 mm; (b) thickness 7,5 mm and (c) thickness 10,0 mm.

Every fracture consists of two stages: (i) crack initiation and (ii) crack propagation. Instrumented Charpy tester can differentiate energies for each stage. Also, it is assumed that every fracture starts with some plastic deformation, no matter how small it is. In the case of ductile fracture, plastic deformation that precedes fracture is considerable. This results in formation of shear lips at the edges of sample. On the other hand, when it comes to brittle fracture, shear lips are not developped, and flat surfaces of non-slip planes in grains are present [11,12].

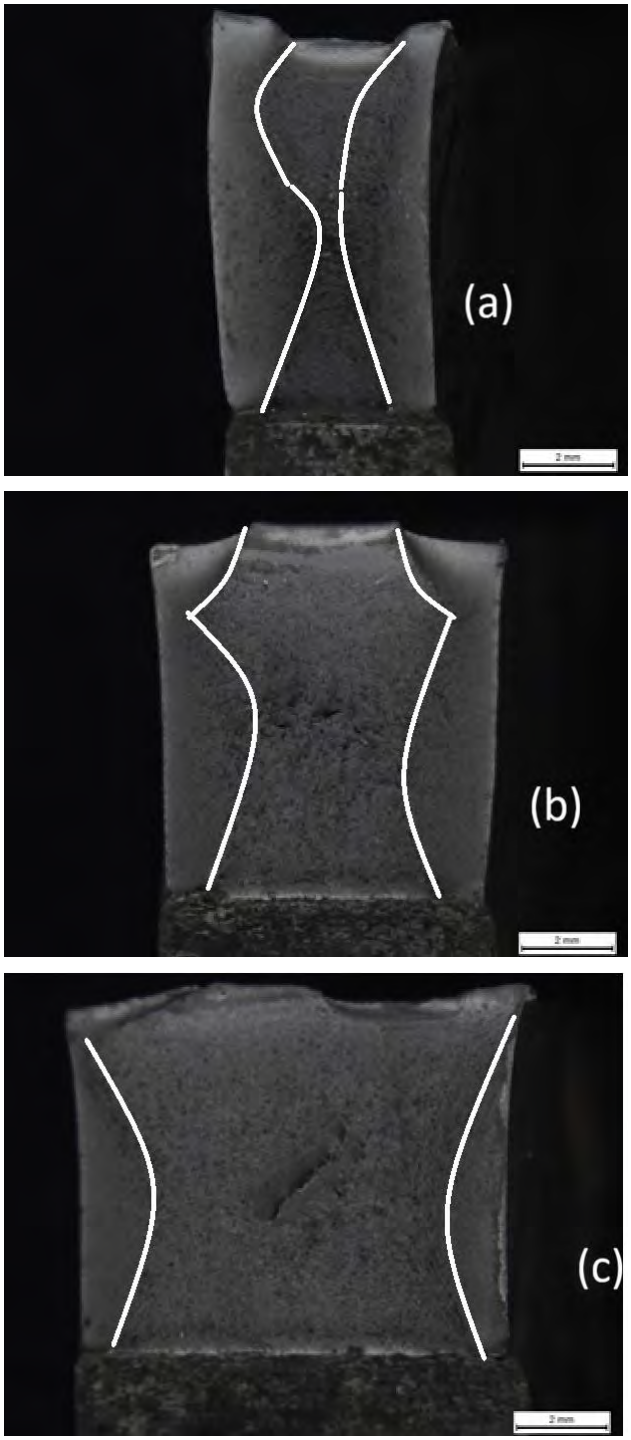


Figure 3. Fractographs of fractured surface of tempered Charpy impact specimens: (a) thickness 5,0 mm; (b) thickness 7,5 mm and (c) thickness 10,0 mm.

First notable feature on fractured surface of quenched and tempered samples is that place of crack initiation can not be unambigously determined, indicating that fracture occured just after attaining critical stress at the tip of crack, which is characteristic for ductile initiation. Also, surface exibits very clear shear lips, that cover substantial surface of cross section. This feature is present in all three samples. Area of shear lips is shown in figure 3 and

Table 3. Charpy impact energies

Thickness t, mm	Quenched and tempered, CVN, J	
	measured	Area normalized
5,0	34.4	34.4
7,5	56,2	51.6
10,0	69,1	68.8

annotated with white lines. Presence of shear lips is proof of ductile mechanism. Shear lips themselves are very pronounced and have position (slope) close to 45˚, confirming presence of plastic deformation that precedes fracture of sample. Interestingly, shear lips look very simillar in all three quenched and tempered samples [11,12].

In order to compare values of impact toughness, measured values were "normalized" by cross section areas. Measured values for 5 mm thick samples were multiplied by factor 1,5 and 2,0 to get value for 7,5 and 10mm, and given in Tables 2 and 3.

Diagram on Figure 4 shows measured data (closed symbols) and normalized values (open symbols) for quenched samples.

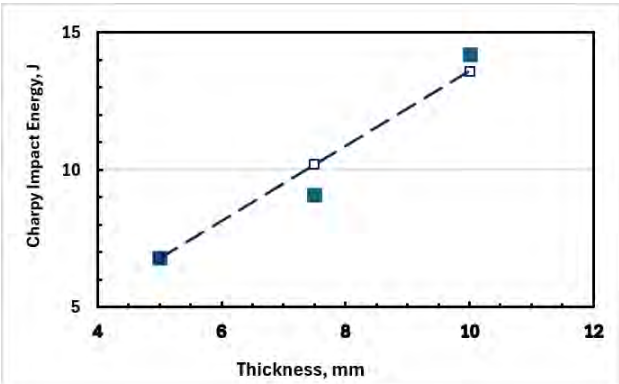


Figure 4. Influence of sample thickness on measured data (closed symbols) and normalized values (open symbols) of Charpy impact toughness for quenched steel.

Data show that there is no linearity and that one value has lower (7,5mm) value, while other is higher (10,0 mm) than expected. It is assumed that this behavior can be attributed dominantly to energy released during plastic deformation. No matter how small it is, due to low toughness (low overall impact energy), its effect is present and can not be neglected. Secondly, not less importantly, impact testing always requires sufficient number of samples to be tested in order to ensure statistical reliability.

Diagram on Figure 5 shows measured data (closed symbols) and normalized values (open symbols) for quenched and tempered samples.

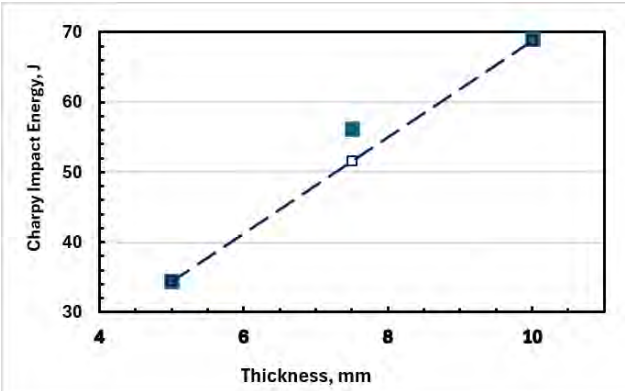


Figure 5. Influence of sample thickness on measured data (closed symbols) and normalized values (open symbols) of Charpy impact toughness for quenched and tempered steel.

Impact toughness in quenched and tempered steels is higher in comparison to quenched samples, as given in tables 2 and 3. This result is in good agreement with previously published data [13-16]. In the quenched sample, crack attain critical length during crack initiation proces, i.e. energy for crack propagation i close to zero. On the other hand, in tempered samples, due to presence of fine particles and grain boundaries, after crack initiation, considerable energy is spent on crack propagation [11-16], due to plastic deformation. Increase in thickness promotes brittle behaviour, but Figure 5 shows that value for 7,5mm is actually some higher than expected.

Data on Figures 4 and 5 show that there is no reliability in potential linking measured impact toughness for one thickness with prediction of its value for different area of cross section.

4. CONCLUSION

An attempt was made to evaluate impact energy as a function of thickness in quenched and tempered low alloyed steel correlate. Steel tested in this work was 42CrMo4 in both quenched and quenched and tempered state. Charpy testing was performed on samples with 5x10 mm2, 7,5x10 mm and 10x10 cross section areas.

In quenched state, cross section is completely flat, indicating brittle fracture and impact energy is small. At the edges of samples, some area that has ductile mechanism is registered. Increase of cross section does not follow any linear law. As a matter a fact, with increase, one value is higher and other smaller than expected. This is attributed to influence of energy absorbed during previous plastic deformation.

In quench and tempered state, impact energy is much higher. Pronounced shear lips on the fractured surface indicate ductile behavior. Their presence is attributed to plastic deformation before critical crack growth.

There is no reliability in potential linking measured impact toughness for one thickness with prediction of its value for different area of cross section, i.e. this prediction is unacceptable.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-34/2026-03/200135) and (Contract No. 451-03-34/2026-03/200325).

References

[1] http://worldsteel.org/ pristupljeno 15.9.26

[2] HOUGARDY, H-P,: Zukünftige Stahlentwicklung. In: Stahl und Eisen 119 (1999) 85 – 90

[3] GALIBOIS,A., KRISHNADEV.M., DUBE´.A.,: Control of grain size and sub-structure in plain carbon and high strength low alloy (HSLA) steels—the problem and the prospect, Metall. Trans. A 10 (1979) 985-995.

[4] DAS.S., CHATTERJEE.S., TARAFDER.S.,: Effect of microstructures on deformation behaviour of high-strength low-alloy steel, J. Mater. Sci. 44 (2009) 1094-1100.

[5] Drobnjak,DJ.,: Physics of Strength and Plasticity, (in Serbian) TMF, Belgrade (1991)

[6] CHARPY, A.G.A. (1901). Essay on the Metals Impact Bend Test of Notched Bars., reprinted in: The Pendulum Impact Testing: A Century of Progress, 2000, ASTM STP 1380, West Conshohoken.

[7] https://www.nde-ed.org/Physics/Materials/Mechanical/Toughness.xhtml, pristupljeno 15.9.26.

[8] https://www.impacttestingmachine.com/new/types-of-impact-testing.html , pristupljeno 15.09.2026.

[9] ASTM E23-02a. Standard Test Methods for Notched Bar Impact Testing of Metallic Materials. ASTM International. Pp 1-27.

[10] EN 10083-3 Steels for quenching and tempering - Technical delivery conditions for alloy steels (2006).

[11] HERTZBERG,R.: Deformation and Fracture Mechanics of Engineering Materials, 4th Edition, John Wiley and Sons, London UK (1995) ISBN: 978-0471012146

[12] BARSOM, J.M. AND ROLFE, S.T.: Fracture and fatigue control in structures : applications of fracture mechanics (1987) Prentice-Hall, Englewood Cliffs, N.J.

[13] PORTER D., EASTERLING.K., AND SHERIF,M.: Phase Transformations in Metals and Alloys, Third Edition CRC Press, Boca Raton, Fl (2009) ISBN: 978-0367430344

[14] Bhadeshia, H.K.D.H., Honeycombe,R.W.K.: Steels: Microstructure and Properties, 4th Edition, Butterworth-Heinemann, London UK (2017) ISBN: 9780081002728

[15] TOTTEN, G.: Steel Heat Treatment – Metallurgy and Technologies, Second Edition, CRC Taylor and Francis, Boca Raton (2007) ISBN: 978-0-8493-8455-4

[16] https://www.phase-trans.msm.cam.ac.uk/2004/Tempered.Martensite/tempered.martensite.html, pristupljeno 16.09.2026

TEH 2026

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES

SECTION VIII

Quality, standardization, metrology, maintenance and exploitation - QSMME

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

FINANCIAL SUSTAINABILITY OF THE DEFENSE SYSTEM OF THE REPUBLIC OF SERBIA: BUDGETARY AND ACCOUNTING ASPECTS

ALEKSANDAR SAVIĆ
Materiel Resources Sector, Ministry of Defense of the Republic of Serbia, Belgrade, aleksandar22071993@gmail.com

MILAN MIHAJLOVIĆ
Military Academy, University of Defense of the Republic of Serbia, Belgrade, milan.mih83@gmail.com

ALEKSANDAR RAJKOVIĆ
Military Academy, University of Defense of the Republic of Serbia, Belgrade, rajkovic.aleksandar1991@gmail.com

Abstract: *The financial sustainability of the defense system is one of the key conditions for the stable functioning of the state and the achievement of its security goals. In conditions of limited budgetary resources and increased demands for the modernization of the armed forces, efficient financial management and a reliable accounting system become necessary for the rational use of public funds. The subject of this paper is the analysis of the financial sustainability of the defense system of the Republic of Serbia with special emphasis on the budgetary and accounting aspects of financing. The aim of the paper is to examine the basic determinants of the financial sustainability of the defense system, as well as the role of budgetary planning, accounting recording and financial reporting in ensuring transparent and efficient management of financial resources. The methodological approach is based on the application of descriptive, analytical and comparative methods, using relevant scientific and professional literature, as well as available statistical and budgetary data. The analysis includes the role of financial management and accounting procedures within the Ministry of Defense of the Republic of Serbia and the financial functioning of the Serbian Armed Forces. The research results indicate the importance of establishing a stable and efficient system of budgetary planning, transparent accounting reporting and strong financial control mechanisms as key factors in the long-term financial sustainability of the defense system.*

Keywords: *defense system, financial sustainability, budget financing, accounting records.*

1. INTRODUCTION

The financial sustainability of the defense system is one of the central issues of modern public finance management, especially in the context of pronounced security challenges and limited fiscal capabilities of the state. The defense system, as a specific and complex component of the public sector, requires the existence of stable, predictable and long-term sustainable sources of financing, while ensuring efficient, rational and responsible management of available financial resources. In this regard, the processes of budget planning, accounting recording and financial reporting are of particular importance, which enable transparent monitoring and control of the spending of public funds. Of particular importance is the analysis of the manner in which these processes are implemented within the Ministry of Defense of the Republic of Serbia, as well as their impact on the overall functioning of the Serbian Armed Forces.

The subject of research in this paper relates to the financial sustainability of the defense system of the Republic of Serbia, with special emphasis on the budgetary and accounting dimensions of financial management. Within this framework, the research includes a detailed consideration of the process of planning, allocation, recording and control of budgetary funds intended for the defense system, while simultaneously analyzing their compliance with the basic principles of efficiency, economy and transparency in the use of public resources.

The research problem arises from the need to determine to what extent the existing system of financial and accounting management contributes to the preservation and improvement of the long-term financial sustainability of the defense system. In this context, the issue of the adequacy and development of the budget planning and accounting system, i.e. their ability to respond to contemporary challenges that include the needs of modernization, continuous growth of costs, as well as limitations arising from the overall state of public finances, is particularly considered.

The main objective of the paper is a comprehensive analysis of the financial sustainability of the defense system of the Republic of Serbia from the perspective of budgetary and accounting management. In addition, specific research objectives have been defined, which include the identification of key determinants of the financial sustainability of the defense system, analysis of the budget planning and execution process, examination of the role of the accounting system in recording and

Corresponding author: A. Savić;
E-mail: aleksandar22071993@gmail.com
DOI: 10.5937/OTEH26140S

controlling budgetary funds, assessment of the efficiency of financial resource management, recognition of the main challenges and limitations in financing the defense system, as well as the formulation of proposals for measures to improve financial sustainability and the accounting system.

Starting from the above objectives, the paper defines a general hypothesis according to which the financial sustainability of the defense system of the Republic of Serbia is directly conditioned by efficient budgetary and accounting management of financial resources. Based on the general hypothesis thus set, specific hypotheses were formulated:
H1: Efficient budget planning and adequate expenditure control significantly contribute to preserving the financial sustainability of the defense system.
H2: A developed, reliable and transparent accounting system has a positive impact on improving financial management in the defense system.
H3: Existing models of financing the defense system face significant limitations arising from the fiscal capacities of the state.
H4: Improving the financial control and audit system contributes to a more rational and responsible use of budget funds.

In order to gain a more complete understanding of the subject matter, the paper applies a combined methodological approach that includes several complementary scientific methods. The descriptive method is used to define the basic concepts and theoretical starting points of financial sustainability and financing of the defense system. The analytical method is applied in considering the budgetary and accounting aspects of financial management, as well as in analyzing the structure of expenditures and financial flows. The comparative method allows for a comparison of different approaches to financing the defense system, as well as the identification of relevant good practices. The content analysis method is used in the study of current legal and by-laws, strategic documents and financial reports. In addition, the statistical method is applied to process and interpret available data on the budget and expenditures of the defense system. The integrated application of the above methods allows for a comprehensive understanding of the financial sustainability of the defense system of the Republic of Serbia, as well as the identification of key factors that influence its improvement.

2. BUDGETARY AND ACCOUNTING FRAMEWORK FOR FINANCING THE DEFENSE SYSTEM

Financial sustainability in the public sector can be defined as the ability of a state to ensure the financing of its functions and the execution of its obligations in the long term, without jeopardizing fiscal stability and overall macroeconomic balance. It implies the existence of a balance between public revenues and expenditures, adequate control of the level of public debt, as well as efficient and rational management of budgetary funds (Schick, 2005). In modern conditions, this concept goes beyond a narrowly understood fiscal discipline and encompasses a broader framework that includes the quality of public spending, the level of transparency and the degree of accountability in the management of public finances (IMF, 2018). Within the defense sector, financial sustainability gains an additional dimension, as it refers to the state's ability to ensure continuous and stable financing of key security functions. This includes financing operational activities, the modernization process, and the development of military capabilities. Sustainable defense financing requires long-term planning, the existence of stable and predictable sources of financing, but also preserving the overall fiscal balance of the state (Hartley, 2011).

The financing of the defense system differs from the financing of other areas of the public sector primarily due to its strategic importance, high level of confidentiality and significant share of capital investments (Knežević et al., 2021). Defense expenditures include a wide range of costs, including current expenditures such as employee salaries, system maintenance and logistical support, as well as capital investments such as the acquisition of weapons and military equipment. These investments are usually long-term and require significant financial resources (Dunne & Perlo-Freeman, 2003). One of the key characteristics of defense financing is the fact that investments in this area do not generate a direct market return. Therefore, the efficiency of these expenditures is not measured through profitability, but through the level of achieved security and the ability to deter potential threats. In addition, defense financing is often strongly influenced by geopolitical circumstances and changes in the security environment, which can lead to significant and sometimes sudden changes in the amount of budget allocations (SIPRI, 2023a).

Budget planning in the defense sector is based on the principles of program budgeting, which provides a clear link between available financial resources, defined objectives and expected results. This approach contributes to improving transparency and increasing efficiency in the allocation of budgetary resources (Robinson, 2013). Within the Ministry of Defense of the Republic of Serbia, the budget process includes the phases of planning, execution and budget control, in accordance with the applicable normative framework, primarily the Law on the Budget System (Milunović et al., 2022). Financial management in this context implies the rational use of funds, continuous control of expenditures and monitoring the implementation of planned activities and goals. A particularly important element is medium-term planning, which enables more stable financing of capital projects and the modernization process of the Serbian Armed Forces (Stevanović et al., 2023). In this way, the risk of inefficient spending of budget funds is reduced and the long-term sustainability of the defense system is ensured.

The institutional framework for financing the defense system in the Republic of Serbia is based on a clearly divided competence between relevant state bodies. The central role is played by the Ministry of Defense of the Republic of Serbia, which is responsible for planning, managing and controlling financial resources allocated to

defense. In addition to this ministry, other actors, such as the Ministry of Finance, the State Audit Institution and various control bodies, play an important role, contributing to ensuring legality, transparency and accountability in the spending of budget funds (Savić et al., 2024a). The normative framework consists of laws and by-laws that regulate the budget system, public procurement and financial operations in the public sector. This institutional setting enables the alignment of strategic defense goals with the general fiscal policy of the state, thus creating conditions for sustainable and stable financing of the defense system. Control and audit play a key role in ensuring the legal, proper and efficient use of public funds. Within the defense system, internal control, internal audit and external audit are distinguished as interconnected oversight mechanisms. Internal control includes a set of procedures, rules and activities that ensure proper financial management and minimize the risk of irregularities, while internal audit is tasked with assessing the efficiency of operations and their compliance with applicable regulations (INTOSAI, 2019a). External audit in the Republic of Serbia is carried out by the State Audit Institution, which provides an independent and objective assessment of the financial operations of budget users. An effectively established control and audit system contributes to reducing the risk of abuse, increases transparency and strengthens public confidence in the functioning of the defense system.

The accounting system in the public sector is based on the principles of budget accounting, which includes recording, classifying and reporting on all financial transactions. In the defense system, accounting plays an extremely important role in monitoring budget execution and providing reliable and relevant information necessary for making management decisions. Financial reporting involves the preparation of periodic and annual reports on budget execution, which contain detailed data on revenues, expenditures and achieved financial results. These reports represent the basis for control, analysis and evaluation of financial operations (IPSASB, 2021). Within the Ministry of Defense of the Republic of Serbia and the Serbian Armed Forces, improving the accounting system, as well as the application of international public sector financial reporting standards (IPSAS), contributes to greater transparency, comparability and reliability of financial information. In this way, the financial sustainability of the defense system is further strengthened and the overall management of public resources is improved.

3. ANALYSIS OF THE FINANCIAL SUSTAINABILITY OF THE DEFENSE SYSTEM OF THE REPUBLIC OF SERBIA

The financial sustainability of the defense system of the Republic of Serbia is conditioned by a number of interrelated factors, among which the stability of budget financing, adequately structured expenditures, optimal level of investment investments, as well as efficient and responsible management of available financial resources stand out. In this part of the paper, attention is focused on the analysis of key elements of defense system financing, with special emphasis on the role of the Ministry of Defense of the Republic of Serbia, as the central institution in the process of planning and managing funds, as well as on the functioning of the Serbian Armed Forces as the main user of budget resources.

Budgetary allocations for defense of the Republic of Serbia in the previous period are characterized by a relatively stable upward trend, which is primarily a result of more complex security circumstances, but also the need for continuous improvement and modernization of the armed forces. The share of defense expenditures in gross domestic product (GDP) ranges from about 1.8% to 2.2%, which is comparable to the practice of most European countries (SIPRI, 2024). A particularly pronounced budget growth was recorded after 2018, when there were more intensive investments in the procurement of weapons and modern military equipment. Such budgetary stability enables more reliable and long-term expenditure planning, which at the same time reduces the risk of disrupting the fiscal balance and ensures the sustainability of the defense financing system (Ministry of Defense of the Republic of Serbia, 2024).

The structure of expenditures is one of the most important indicators of financial sustainability, since it indicates the way in which funds are distributed between current needs and development priorities. Within the defense system of the Republic of Serbia, three basic categories of expenditures are dominant: expenditures for employees, operating costs and capital expenditures, i.e. investments in equipment and infrastructure. The following paper presents the framework structure of defense system expenditures.

Table 1. Structure of defense system expenditures of the Republic of Serbia in the average amount for the period 2018-2025.

Expenditure category	Share (%)
Employee expenses	45–50%
Operating costs	25–30%
Investments and equipment	20–25%

Source: Prepared by the author based on data from the Ministry of Defense and the Stockholm International Peace Research Institute (SIPRI, 2024)

The relatively high share of personnel expenditures indicates a pronounced personnel component of the defense system, while the growth of capital expenditures in recent years clearly indicates an increased orientation towards modernization and technological improvement (SIPRI, 2024).

Investments represent a key pillar of the long-term financial sustainability of the defense system. These investments cover a wide range of activities, including the procurement of modern weapons and military equipment, the digitalization of management systems, as well as the improvement of infrastructure and logistics capacities (Mihajlović et al., 2023). In the previous period, the

Republic of Serbia has significantly intensified investments in the modernization of combat systems, the development of the domestic military industry and the strengthening of logistical support. Such investments have a double effect: on the one hand, they contribute to increasing the operational capabilities of the defense system, while on the other hand, they create additional pressure on the state budget. For this reason, it is of particular importance to establish a balance between investment needs and preserving fiscal stability (Petrović, 2021).

The efficiency of financial resource management is reflected in the ability of the defense system to achieve maximum results in terms of functionality and operational readiness with limited resources (Savić et al., 2025a). Key factors that influence the level of efficiency include the quality of budget planning, the application of the principles of program budgeting, the development of the accounting system, as well as the efficiency of internal and external control. The introduction of modern accounting standards and the improvement of the financial reporting system significantly contribute to greater transparency and accountability in the management of public funds (Đorđević, 2023). Also, the development of internal audit and financial control systems enables timely detection of irregularities and improvement of the resource management process. However, despite the progress made, certain challenges remain, especially in the areas of expenditure optimization, long-term investment planning and harmonization of the strategic goals of the defense system with the realistic fiscal constraints of the state. Efficient and responsible financial management remains a key prerequisite for ensuring stability and long-term financial sustainability of the defense system of the Republic of Serbia.

4. CHALLENGES TO THE FINANCIAL SUSTAINABILITY OF THE DEFENSE SYSTEM

The financial sustainability of the defense system is a complex and multi-layered concept that depends on a combination of internal resource management factors and external economic and security circumstances. Within the Ministry of Defense of the Republic of Serbia and the Serbian Armed Forces, these sustainability challenges are particularly evident in situations of limited public funds, continuous growth in modernization costs, and an increasingly dynamic and uncertain security environment (Savić et al., 2023).

One of the key challenges to the financial sustainability of the defense system relates to the limitations of the state budget. The defense budget represents a part of total public expenditures and largely depends on macroeconomic stability, the level of public debt, and the general fiscal policy of the state (Savić & Bonić, 2022). Given that the state must simultaneously provide financing for numerous sectors such as health, education, and social protection, defense allocations are often under pressure to be rationalized (Mowery, 2010). The Republic of Serbia, as a state with limited fiscal capacities, is faced with the need to align the real needs of the defense system with available budgetary opportunities. According to the Stockholm International Peace Research Institute, defense spending in developing countries is not stable but varies depending on economic cycles and political priorities (SIPRI, 2023b), which further indicates the importance of long-term and stable financial planning in the field of defense.

Modern defense systems require constant and significant investments in modernization and technological development. The costs of purchasing modern combat systems, equipment and supporting infrastructure represent a major burden on the defense budget. The modernization process of the Serbian Armed Forces involves the procurement of complex and financially demanding systems, the total costs of which often exceed the capabilities of short-term budget planning (Hartley, 2012). An additional challenge is reflected in the fact that defense investments are long-term in nature, with their effects not immediately visible, while expenditures are realized in a relatively short period of time. Such an asymmetry between investments and effects creates additional pressure on fiscal stability and requires careful and responsible planning of capital expenditures.

The efficiency of financial management in the defense system largely depends on the institutional framework and organizational structure. Within the Ministry of Defense of the Republic of Serbia, challenges relate to the need to improve coordination between different organizational units, as well as to strengthen the internal control and internal audit systems. Insufficiently developed accounting systems and limited transparency of financial reporting can negatively affect the efficiency of public fund management (Allen & Tommasi, 2001). In addition, the implementation of modern international standards in the public sector, including IPSAS, continues to be a challenge for a large number of countries in transition, including for the improvement of the financial accounting system in the field of defense (Young, 2018). The security environment has a direct and strong impact on the volume and structure of defense financing (Stojković & Glišić, 2020). Global geopolitical changes, regional tensions, and modern forms of threats, such as hybrid and cyber threats, require increased investments in defense capacities (Smith, 2020). In such circumstances, states are often forced to increase budget allocations for defense, which further burdens public finances. At the same time, the high degree of uncertainty regarding future security challenges makes long-term planning and optimal allocation of resources difficult.

Table 2. Budget allocation for the needs of the defense system of the Republic of Serbia

Year	Defense Budget (billion RSD)	% GDP
2018	70	1.6%
2019	80	1.7%
2020	95	1.9%

2021	102	2.0%
2022	120	2.1%
2023	138	2.2%
2024	158	2.3%
2025	261	3.5%

Source: Prepared by the author based on data from the Ministry of Defense and the Stockholm International Peace Research Institute (SIPRI, 2024)

The table shown indicates a clear trend of increasing defense spending, which on the one hand reflects the increased needs for financing the defense system, while on the other hand raises the question of long-term fiscal sustainability and the ability of the state budget to continuously support such growth.

5. POSSIBILITIES FOR IMPROVING THE FINANCIAL SUSTAINABILITY OF THE DEFENSE SYSTEM

The financial sustainability of the defense system of the Republic of Serbia requires continuous improvement of public resource management, with consistent application of modern budgetary, accounting and control instruments. Within the Ministry of Defense of the Republic of Serbia and the Serbian Armed Forces, improving financial sustainability implies a comprehensive and systemic approach that includes improving the budget planning process, strengthening the accounting system and financial reporting, developing financial control mechanisms, as well as adopting and implementing internationally recognized good practices in the field of public finance (Radivojević et al., 2022).

Budget planning is the main instrument used to ensure the financial stability of the defense system. The traditional budget planning model, based on an incremental approach, is often not fully adequate for modern challenges, and there is a need to improve it through the introduction of program and medium-term budgeting. Program budgeting allows for direct linking of expenditures with clearly defined objectives and expected results, which significantly improves transparency and efficiency in the use of public funds (Robinson, 2015). On the other hand, the Medium-Term Expenditure Framework (MTEF) contributes to more stable and predictable resource management, as well as better alignment of strategic defense objectives with the fiscal capabilities of the state (OECD, 2019). This approach is of particular importance in the defense sector, given that investments in weapons, equipment, and infrastructure require multi-year and continuous financial planning. Further improvement of budget planning also includes the development of analytical capacities for cost assessment, as well as the wider application of analytical methods such as cost-benefit and cost-effectiveness analysis, which contributes to more rational decision-making on the allocation of limited budgetary resources (Allen, Hemming, & Potter, 2013). An efficient accounting system is one of the key components of financial sustainability, as it enables accurate, timely, and transparent recording of all financial flows (Savić et al., 2024b). In the modern public sector, and therefore in the defense system, the transition from cash to accrual accounting is of increasing importance, in accordance with international public sector standards, such as IPSAS (International Public Sector Accounting Standards) (IFAC, 2021). The application of accrual accounting allows for a more realistic and comprehensive presentation of the financial position, including liabilities, assets and long-term costs, which is especially important for the defense system that manages extensive and valuable tangible assets. Transparent financial reporting contributes to strengthening public trust, as well as improving fiscal discipline and accountability in the management of public funds (Chan, 2010). Within the Ministry of Defense of the Republic of Serbia, improving the accounting system involves further digitalization of financial processes, integration of information systems and standardization of financial reporting, which enables more efficient resource management and better management decision-making (Savić et al., 2025b).

Financial control is an important mechanism that ensures the legal, economical and efficient use of budgetary funds. In the defense system, the importance of financial control is further emphasized due to the complexity and volume of financial flows and operations. The modern concept of financial control involves the development of an internal control and internal audit system, based on the principles of the COSO framework (Committee of Sponsoring Organizations of the Treadway Commission, 2017). The internal control system includes risk management processes, control activities and continuous monitoring, which reduces the risk of irregularities, abuse and inefficient spending of funds. In addition to internal control, external audit also plays an important role, contributing to a greater degree of transparency and accountability in the management of public finances (INTOSAI, 2019b). Strengthening cooperation between internal and external control mechanisms can significantly contribute to improving the overall system of financial supervision and control.

The application of international good practices is one of the key directions for improving the financial sustainability of the defense system. International organizations, such as the OECD and NATO, have developed recommendations and standards related to transparency, accountability and efficiency in defense budget management (OECD, 2019). One of the basic principles is the transparency of the defense budget, which implies the availability and clarity of data on the planning and execution of budget funds. In addition, the concept of "value for money", which refers to the optimal use of available resources while achieving maximum effects and results, is particularly important (Allen et al., 2013).

The experiences of developed countries show that integrated information systems, professionalization of personnel in the field of finance and accounting, as well as continuous education of employees, are key factors in

efficient management of public finances. In this regard, the defense system of the Republic of Serbia can further improve its capacities through the adaptation and implementation of the above practices, while harmonizing them with national institutional and fiscal specificities.

6. CONCLUSION

The financial sustainability of the defense system is a complex and multi-layered concept that involves harmonizing the security needs of the state with its available fiscal capacities. This paper analyzes the financial sustainability of the defense system of the Republic of Serbia, with a special focus on budgetary and accounting aspects. Based on the analysis conducted, it can be concluded that stable, responsible and efficient management of public finances is a basic prerequisite for the long-term and smooth functioning of the defense system.

The results obtained indicate that the level of financial sustainability of the defense system largely depends on the quality of the budget planning process, the reality in the assessment of necessary expenditures, as well as the consistent application of the principles of fiscal discipline. In this sense, the Ministry of Defense of the Republic of Serbia has a key role in establishing and implementing a system of planning, distribution and control of budgetary funds, while the Serbian Armed Forces are the main user of these funds, whose operational capability and efficiency largely depend on the rational and purpose-oriented spending of allocated resources.

An important role in ensuring financial sustainability is played by the accounting system, which enables precise recording of all financial flows, transparent financial reporting, and the creation of a reliable basis for making management decisions. The paper emphasizes that improving accounting procedures and the quality of financial reporting directly contributes to increasing transparency, strengthening accountability, and better control over the use of budget funds, thereby improving the overall efficiency of resource management and reducing the risk of irrational spending.

The analysis also showed that the key challenges in the area of financial sustainability of the defense system relate to the limited budget funds, the increasing costs of modernization and equipment, and the need for continuous improvement of material and technical capabilities. In addition, the dynamic and changing security environment imposes the need for a more flexible, but at the same time long-term sustainable approach to expenditure planning, while preserving the overall fiscal stability of the state.

In order to improve the financial sustainability of the defense system of the Republic of Serbia, it is necessary to further develop the program approach to budgeting, strengthen financial control and internal audit mechanisms, and improve the accounting framework in accordance with modern international public sector standards. Also, special importance should be given to long-term planning of investments and modernization processes, with clear definition of priorities and ensuring optimal allocation of available resources.

Finally, it can be concluded that the financial sustainability of the defense system is not a static category, but a continuous and dynamic process that requires constant adaptation to economic and security changes. Only through the coordinated operation of budget planning, financial management and accounting systems is it possible to ensure a stable, efficient and long-term sustainable defense system of the Republic of Serbia.

References

[1] Allen, R., & Tommasi, D. (2001). Managing public expenditure: A reference book for transition countries. OECD.

[2] Allen, R., Hemming, R., & Potter, B. (2013). *The international handbook of public financial management*. Palgrave Macmillan.

[3] Chan, J. L. (2010). IPSAS and government accounting reform in developing countries. *Accounting and Business Research*, 40(3), 123–135.

[4] Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2017). *Enterprise risk management: Integrating with strategy and performance*. COSO.

[5] Đorđević, M. (2023). *Finansijsko upravljanje u javnom sektoru*. Beograd: Ekonomski fakultet.

[6] Državna revizorska institucija. (2022). Izveštaji o reviziji budžetskih korisnika u Republici Srbiji.

[7] Dunne, J. P., & Perlo-Freeman, S. (2003). The demand for military spending in developing countries. *International Review of Applied Economics*, 17(1), 23–48. https://doi.org/10.1080/713673166

[8] Hartley, K. (2011). The economics of defence policy: A new perspective. Routledge.

[9] Hartley, K. (2012). The economics of defence policy: A new perspective. Routledge.

[10] IFAC. (2021). Handbook of International Public Sector Accounting Pronouncements. International Federation of Accountants.

[11] International Monetary Fund (IMF). (2018). *Fiscal transparency handbook*. IMF.

[12] INTOSAI. (2019a). Guidelines for internal control standards for the public sector.

[13] INTOSAI. (2019b). *Principles of public-sector auditing*. International Organization of Supreme Audit Institutions.

[14] IPSASB. (2021). Handbook of International Public Sector Accounting Pronouncements.

[15] Knežević, M., Nikolić, S., & Neševski, A. (2021). The budget control in the function of the civil and democratic control of the armed forces in the Republic of Serbia. *Vojno delo*, *73*(3), 103-120. https://doi.org/10.5937/vojdelo2103103K

[16] Mihajlović, M., Peković, J., & Veljković, I. (2023).

The comparative analysis of the defence system costs of the Republic of Serbia and the NATO member states. *Vojno delo*, *75*(3), 13-29. https://doi.org/10.5937/vojdelo2303013M

[17] Milunović, M., Knežević, M., & Kovačević, N. (2022). The improvement of the financial management and control in the Ministry of Defence of the Republic of Serbia. *Vojno delo*, *74*(5), 42-56. https://doi.org/10.5937/vojdelo2205042M

[18] Ministarstvo odbrane Republike Srbije. (2018-2024). *Izveštaji o budžetu i finansiranju odbrane*. Beograd.

[19] Mowery, D. C. (2010). Military R&D and innovation. *Handbook of the Economics of Innovation*, 2, 1219–1256.

[20] OECD. (2019). Budgeting and public expenditures in OECD countries. OECD Publishing.

[21] OECD. (2023). *Government at a Glance*. Paris: OECD Publishing.

[22] Petrović, P. (2021). *Javne finansije i fiskalna politika*. Beograd: CID.

[23] Radivojević, M., Tešanović, B., & Majstorović, M. (2022). Budget Financing of the Defense System and Opportunities for Improvement. *International journal of economics and law*, 191.

[24] Robinson, M. (2013). Program budgeting and performance-based budgeting. IMF.

[25] Robinson, M. (2015). From old public administration to the new public service: Implications for public sector reform. World Bank.

[26] Savić, A. & Bonić, Lj. (2022). Analysis of the impact of reporting on environmental performance idicator on the profitability of European companies, Facta Universitatis – Economics and Organization, 19 (3), 167-182. https://doi.org/10.22190/FUEO220529013S

[27] Savić, A. & Mihajlović, M. (2025b). Quantitative evaluation of the application of the code of ethics for accountants in business entities. Časopis Ekonomist, 4 (2), 57-92. https://doi.ub.kg.ac.rs/doi/casopisi/10-46793-ekonomist4-2-3-s/

[28] Savić, A. O., Mihajlović, M. M., & Božović, I. D. (2023). Macroeconomic aspects of comprehensive costs of assets as a prerequisite for equipping the defense system with weapons and military equipment. *Vojnotehnicki glasnik/Military Technical Courier*, *71*(3), 797-815. https://doi.org/10.5937/vojtehg71-44145

[29] Savić, A., Mičić, S, Ristić, D. & Radosavljević, M. (2024a). Kvantitativna analiza revizije svrsishodnosti nabavki korisnika javnih sredstava. Časopis DIT., 42, 115-122. http://diz.org.rs/images/casopis/dit42.pdf

[30] Savić, A., Mihajlovic, M., & Kostić, R. (2025a). Reconciliation of amounts in budget execution reports and financial statements using the example of the united nations'food and agriculture organization (FAO). *Economic of Agriculture*, *72*(4), 1163-1183. https://doi.org/10.59267/ekoPolj25041163S

[31] Savić, A., Ranisavljević, M. & Mihajlović, M. (2024b). The role of intellectual capital reporting in the process of supply chain integration of the military technical institute. Scientific Technical Review, 74 (2), 71-76. https://doi.org/10.5937/OTEH2402071S

[32] Schick, A. (2005). Sustainable public sector finances. *OECD Journal on Budgeting*, 5(1), 107–126.

[33] SIPRI. (2023a). SIPRI Yearbook 2023: Armaments, disarmament and international security.

[34] SIPRI. (2023b). *SIPRI Military Expenditure Database*. Stockholm International Peace Research Institute.

[35] SIPRI. (2024). *SIPRI Military Expenditure Database*. Stockholm: Stockholm International Peace Research Institute.

[36] Smith, R. (2020). Defence economics and security challenges. *Defence and Peace Economics*, 31(2), 123–135.

[37] Stevanović, T., Antić, L., & Savić, A. (2019). Specificity of performance measurement in the ministry of defense and the Serbian armed forces. *Facta universitatis, Series: Economics and Organization*, 283-297. https://doi.org/10.22190/FUEO1903283S

[38] Stojković, D., & Glišić, M. (2020). Serbia's military neutrality: Is it economically beneficial?. *Defence and Peace Economics*, *31*(5), 583-599. https://doi.org/10.1080/10242694.2018.1547952

[39] Vlada Republike Srbije. (2023). Zakon o budžetu Republike Srbije za 2018-2024. godinu. Beograd.

[40] Young, T. D. (2018). The failure of defense planning in European Post-Communist Defense Institutions: ascertaining causation and determining solutions. *Journal of Strategic Studies*, *41*(7), 1031-1057. https://doi.org/10.1080/01402390.2017.1307743

[41] Zakon o budžetskom sistemu Republike Srbije („Službeni glasnik RS", br. 54/2009, 73/2010, 101/2010, 101/2011, 93/2012, 62/2013, 63/2013 - ispr., 108/2013, 142/2014, 68/2015 - dr. zakon, 103/2015, 99/2016, 113/2017, 95/2018, 31/2019, 72/2019, 149/2020, 118/2021, 138/2022, 118/2021 - dr. zakon, 92/2023 i 94/2024)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

VARIABILITY OF BALLISTIC TEST RESULTS AND THE RELIABILITY OF QUALITY ASSESSMENT OF PERSONAL BALLISTIC PROTECTION EQUIPMENT

MARKO MILOVANOVIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, marko.milovanovic547@gmail.com

SINIŠA GAĆIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, gacic.sinisa@gmail.com

NIKOLA PRODANOVIĆ
Security Information Agency, Laboratory for testing of protective equipment, Belgrade, nikola.prodan7@gmail.com

Abstract: *Ballistic testing represents the definitive method for quality control of personal ballistic protective equipment. Testing procedures are defined by relevant standards, yet in practice, deviations in results can occur, raising questions about the reliability of equipment quality assessments. This study analyzes results obtained from testing samples of personal ballistic protective equipment in three accredited European laboratories according to NIJ 0101.04 and STANAG 2920 standards. The comparison includes measurements of backface signature (BFS) and the V_{50} parameter, aiming to determine the extent of deviations among obtained values. The analysis considers factors that may influence observed differences, including the properties of the backing material (Roma Plastilina), projectile velocity measurement conditions, and the procedure for determining the V_{50} parameter. The results indicate that the observed deviations can be significant enough to affect the interpretation of test outcomes in quality control procedures. The study highlights the importance of systematically considering result variability and provides a basis for improving procedures for assessing the reliability of ballistic testing, contributing to a better understanding of the limitations and practical application of these methods in the quality control of personal ballistic protective equipment.*

Keywords: *clay, backface signature, probit method, inter-laboratory testing.*

1. INTRODUCTION

Ballistic testing represents the definitive method for the quality control of personal ballistic protective equipment. Although testing procedures are strictly defined by international standards such as NIJ 0101.04 and STANAG 2920, significant deviations in results can occur in practice, raising fundamental questions about the reliability of protective equipment quality assessments.

A primary source of variability in NIJ 0101.04 is the backing material, specifically Roma Plastilina No. 1 (RP#1) clay. Research [1] has shown that the correlation between projectile impact and Backface Signature (BFS) in this material is often low, whereas studies on human cadavers show a much higher correlation. Furthermore, no formal technical standards exist for the procurement and acceptance testing of RP#1, meaning its consistent quality cannot be guaranteed. Due to its non-Newtonian rheological properties, the mechanical behavior of the clay is highly sensitive to both temperature and its "shear history" or previous working [2]. Previous study [3] has already indicated that different production batches of the same clay can lead to significant variations in BFS depth during mandatory calibration .

This study is significant because it provides an analysis of inter-laboratory testing data, which has been largely unavailable until now, to demonstrate how these factors influence the assessment of ballistic protection standards. By analysing measurements of BFS and the V_{50} parameter across three accredited European laboratories, this paper aims to identify the extent of these deviations and provide a basis for improving the reliability of ballistic testing methodologies.

2. EXPERIMENT

The tests were conducted comparatively across three accredited laboratories:

1. Security Information Agency, Laboratory for Testing of Protective Equipment, Belgrade, Republic of Serbia,
2. Banco Nazionale di Prova per le Armi da Fuoco Portatili e le Munizioni Commerciali, Brescia, Italy,
3. AITEX, Research & Innovation Center, Alcoy, Spain.

For this study, ballistic panels with the following configuration were used, as observed from the strike face:

- 17 layers of UHMWPE Dyneema SB51,
- 1 layer of aramid FMS AT245,

Corresponding author: M. Milovanović;
E-mail: marko.milovanovic547@gmail.com
DOI: 10.5937/OTEH26141M

- 1 layer of UHMWPE Dyneema SB51,
- 1 layer of FMS AT245, and
- 3 layers of UHMWPE Dyneema SB51.

The UHMWPE Dyneema SB51 sheets were from batch number 2200275436, and the aramid FMS AT245 sheets from batch number 7126. Each panel sample measured (400×400) mm. Three samples were provided to each laboratory.

To ensure an unbiased analysis of the results, each participating laboratory was assigned a unique code (A, B, and C).

2.1 Method for determination of backface signature (BFS) - NIJ 0101.04

Ballistic testing was conducted in accordance with the methodology specified for evaluating the ballistic resistance of personal body armor against gunfire, as defined by the NIJ Standard 0101.04.

All shots were fired using a 9×19 mm test barrel, with 9×19 mm ammunition loaded with 8.0 g full metal jacket (FMJ) bullets. The reference velocity of the bullets was (436±9) m/s, in accordance with the standard requirements for protection level IIIA. Testing was performed on dry samples.

Impact locations were defined in Article 5.11.3 of the standard, and the shot layout is shown in Figure 1. Shots numbered 1, 2, 3, and 6 were fired at an angle of 0°, while shots 4 and 5 were fired at a 30° angle relative to the center of the sample.

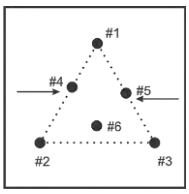


Figure 1. Impact position layout

Each sample was positioned 5 m from the muzzle of the test barrel, and bullet velocity was measured at a distance of 2.5 m from the sample.

2.2 Ballistic test method for personal armour materials and combat clothing - STANAG 2920

The testing was conducted in accordance with STANAG 2920, Edition 2. The sample panel was locked in a special steel frame, consisting of two metal frames that were tightened together. The frame-panel assembly was anchored to a support so that the panel could absorb all the kinetic energy possessed by the FSP at the moment of impact. All tests were performed on dry samples.

Shooting was performed by using a .22 caliber test barrel. Type 2 fragment-simulating projectiles (FSP) were used, with characteristics specified in Appendix 1 of Annex A (Category A2).

Test area ambient conditions complied with the requirements of Section 7.3.

For the velocity measurement all laboratories used optical light gates and measured the FSP velocity at 2.5 m from the sample. Since impact velocity is much lower due to air drag, it was calculated by using some of the special analytical solutions of flat-fire trajectories equations, which give the smallest error [4].

The fragments were fired from a distance of 5 m from the front face of the sample with an angle of incidence of 0°. The distance between the impact point and previous shots or the edge of the test specimen was at least 5 cm.

After each shot, checks were carried out to see whether the fragment had been stopped or had penetrated the witness plate. An aluminium sheet was used as a witness plate, with a thickness of 0.5 mm, and it was placed 150 mm from the rear face of the sample.

During the testing, six FSP impacts were performed, of which three resulted in complete perforations and three in non-perforations. The nominal initial velocity was set at 600 m/s, with a maximum velocity variation range of 40 m/s between the lowest and the highest achieved values. The final result was reported as the arithmetic mean velocity of the three complete perforations and three non-perforations, expressed to one decimal place.

3. RESULTS AND DISCUSSION

3.1 Method for determination of backface signature (BFS) NIJ 0101.04

The test results are presented in Table 1, showing the BFS values and impact velocities for shots 1, 2, and 3. In accordance with the NIJ Standard 0101.04, only these specific shots were considered for evaluation. Although the backing material was calibrated both before and after each test, the values reported in Table 1 refer specifically to the pre-test calibration. Individual shots are identified using Roman numerals. Table 2 provides the descriptive statistics for BFS values, organized by individual sample and aggregated by laboratory. In this table, mean values are denoted as x_{aver}, the range as D, and the standard deviation as σ. All statistical data processing was performed using Minitab v.21.4.2.

To ensure the integrity of the data, Dixon's and Grubb's tests were initially applied to identify potential outliers in BFS and bullet velocity measurements. These tests confirmed that all extreme values remained within permissible limits; specifically, p-values for Dixon's test were 0.553 and 0.162, while Grubb's test yielded p-values of 0.696 and 0.278. Consequently, the sample contains no outliers, indicating that the results are homogeneous and reliable—a finding further supported by the boxplot diagram in Figure 2.

Analysis of Table 1 reveals that BFS values range from 22 mm to 29 mm, resulting in a total spread of 7 mm. For individual samples, the ranges vary between 0 mm and 5 mm, though they typically average 4 mm. A similar trend is observed across laboratories, where ranges are generally 4 mm, with the exception of Laboratory C,

which recorded a range of 6 mm. Fisher's test for equality of variances confirmed that the standard deviations are not statistically significantly different.

Table 1. Test results for the determination of the backface signature

Laboratory code	Sample	Calibration, [mm]	Backface signature, BFS, [mm]			Bullet velocity, [m/s]		
			I	II	III	I	II	III
A	1	18.0	25	25	29	437.8	442.2	435.4
	2	17.4	26	23	28	441.1	435.3	444.2
	3	20.9	27	27	27	439.4	428.6	435.9
B	1	19.2	25	26	23	442.5	445.3	437.0
	2	19.6	23	25	27	437.3	437.8	442.9
	3	19.8	25	26	26	435.5	432.8	442.3
C	1	18.6	23	26	25	437.6	439.2	436.5
	2	18.6	22	26	26	432.6	445.4	440.2
	3	18.6	23	25	23	441.1	439.7	434.6

Table 2. Descriptive statistics on results of BFS (all measures are expressed in mm)

Lab. code	Sample	x_{aver}	D	σ	x_{aver}	D	σ
A	1	24.7	3	1.52	25.1	4	1.36
	2	25.0	4	2.00			
	3	25.7	1	0.33			
B	1	24.7	3	1.53	24.3	4	1.58
	2	24.7	4	2.31			
	3	23.7	2	0.67			
C	1	26.3	4	2.31	26.3	6	1.80
	2	25.7	5	2.50			
	3	27.0	0	0.00			

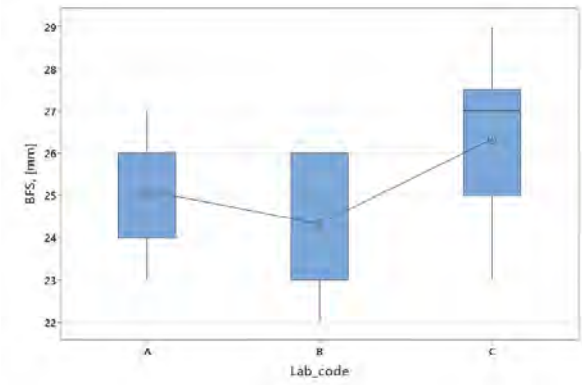


Figure 2. Boxplot and interval plot of BFS values

The boxplot in Figure 2 shows noticeable skewness in the results from Laboratories B and C. This is particularly pronounced for Laboratory C, where the median is significantly higher than those of the other two laboratories. At first glance, this suggests that the test results may vary significantly depending on the laboratory where the testing is conducted. To analyze this numerically, a one-way ANOVA was performed using the laboratory location as the primary factor (Table 3).

Table 3. ANOVA test for BFS

Source	DF	Seq SS	Contrib.	Adj MS	F-value	p-value
Lab	2	18.30	23.11%	9.148	3.61	0.043
Error	24	60.89	76.89%	2.537		
Total	26	79.19	100%			

The ANOVA test produced a p-value slightly below 0.05, indicating a statistically significant difference in test results between laboratories. To identify specific differences, Tukey pairwise comparisons and Hsu's Multiple Comparisons with the Best (MCB) were conducted. Both methods confirmed that the significant differences exist between Laboratories B and C. The model's coefficient of determination (R^2) was 23.11%, while the predicted R^2 was 2.68%, suggesting that the model provides insufficient precision for predicting new observations, likely due to the small group sizes.

Given the limited sample size, evaluating the residual plots to verify ANOVA assumptions is challenging. Therefore, the Ryan-Joiner test was applied to the residuals to confirm normality, alongside Bartlett's and Levene's tests for constant variance. The RJ test statistic was 0.985 ($p > 0.05$), providing insufficient evidence at the α=0.05 significance level to suggest a violation of the normality assumption. Similarly, p-values for Bartlett's and Levene's tests (0.747 and 0.754, respectively) indicate that the assumption of constant variance holds, further validating the initial Fisher's test results.

To determine the underlying causes of the observed BFS ranges, a multivariate regression analysis was conducted to examine the impact of bullet velocity and the initial hardness of the backing material. Additionally, the influence of the laboratory factor on the relationship between BFS, velocity, and calibration was analyzed (Table 4). The Pareto chart in Figure 3 shows that only the laboratory factor exceeds the critical threshold. Of the remaining variables, bullet velocity had a secondary influence, while calibration (initial hardness) had the least impact.

The results in Table 4 show that only the laboratory factor yielded a p-value below 0.05. This leads to the conclusion that BFS does not fundamentally correlate with bullet velocity or backing material calibration. Thus, while the initial hardness of the backing material varied across samples and laboratories, it did not statistically affect the results. This suggests that the behavior of the backing material—due to inherent non-homogeneity, previous test history, and mechanical treatment—is inconsistent across the entire contact surface. This inconsistency in calibration behavior is also evident across different

versions of the NIJ 0101 standard. In Version 3, calibration utilized a 44.5 mm diameter weight with an allowable indentation of (25 ±3) mm. In Version 4, the diameter is 63 mm with an allowable depth of (19 ± 2) mm, using the same 1 kg mass dropped from 2 meters.

Table 4. ANOVA test for regression analysis

Source	DF	Seq SS	Contrib.	Adj SS	F-value	p-value
Regression	4	26.1	32.98%	26.1	2.71	0.057
Velocity	1	3.83	4.84%	6.98	2.90	0.103
Calibration	1	0.16	0.20%	2.09	0.87	0.362
Lab	2	22.1	27.94%	22.1	4.59	0.022
Error	22	53.1	67.02%	53.1		
Total	26	79.19	100%			

Furthermore, the interaction between the sample and the backing material must be considered, particularly the energy transfer mechanism. Factors such as the mounting method, the rigidity of the attachment to the support, and any movement during testing represent laboratory-specific variables that must be minimized. Additionally, the operator's influence during BFS measurement is a factor, separate from the quantifiable uncertainty of the measuring instrument itself.

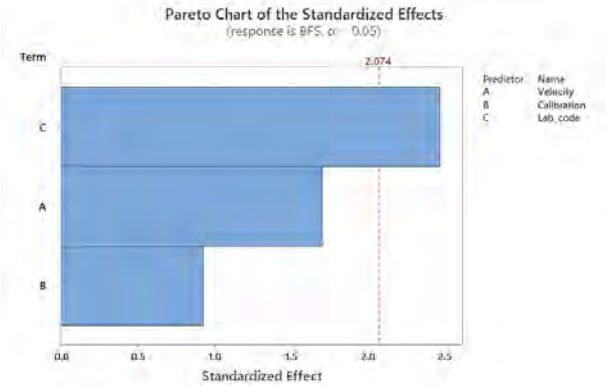


Figure 3. Pareto chart of the standardized effects

The reasoning for the laboratory's influence on the regression analysis is visible in the interaction plot for BFS (Figure 4). In Laboratory C, BFS remains virtually unchanged as bullet velocity increases. In contrast, Laboratory B shows the most significant increase in BFS relative to velocity. To further analyze this, a univariate regression analysis was conducted for each laboratory, focusing solely on velocity since backing material calibration results were proven non-influential (Table 5).

For Laboratory A, the R^2 was 0.4398 (Adjusted R^2 = 0.3598), meaning 43.98% of the variance is explained by the model, while the remaining 56.02% is attributed to random error. The linear regression graph for Laboratory A is shown in Figure 5. For Laboratory B, the R^2 was 0.1655 (Adjusted R^2 = 0.0463), significantly lower than Laboratory A. In Laboratory C, the R^2 was 0.0001, indicating that 99.99% of the model is random error.

Residual analysis for each laboratory confirmed that residuals were randomly and evenly distributed, satisfying the requirements for normality and lack of bias, with no residual exceeding two standard deviations. Since p-values for all three laboratories (especially Laboratory C) were greater than 0.05, it is confirmed that any apparent correlation between velocity and BFS is not statistically grounded at the 5% significance level. This conclusion implies that, when following the NIJ 0101.04 standard, a higher velocity in shots 2 or 3 does not necessarily result in a deeper indentation in backing material.

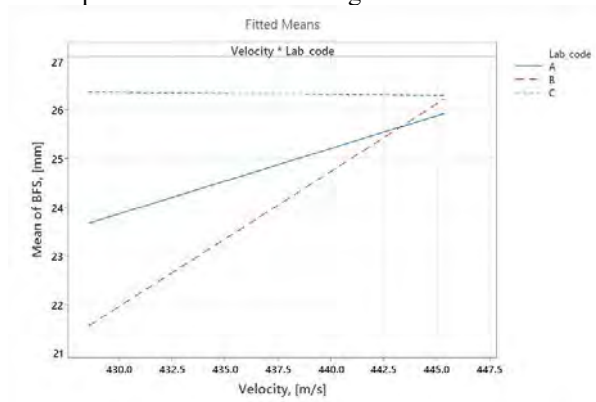


Figure 4. Interaction plot for BFS

Table 5. ANOVA test for univariate regression analysis

Lab. code	Source	DF	SS	MS	F-value	p-value
A	Reg.	1	8.80	8.80	5.50	0.052
	Error	7	11.2	1.60		
	Total	8	20.0			
B	Reg.	1	2.46	2.46	1.39	0.277
	Error	7	12.4	1.78		
	Total	8	14.9			
C	Reg.	1	0.00	0.00	0.00	0.979
	Error	7	26.0	3.71		
	Total	8	26.0			

While laboratories report measurement uncertainty regarding the depth measurement itself, the uncertainty of the method—which includes the backing material's characteristics—is often overlooked. The results of this study demonstrate that BFS variability is significantly higher than the measurement uncertainty of the instrument alone. This discrepancy raises questions about the reliability of quality assessments when results are near the maximum allowable limit of 44 mm.

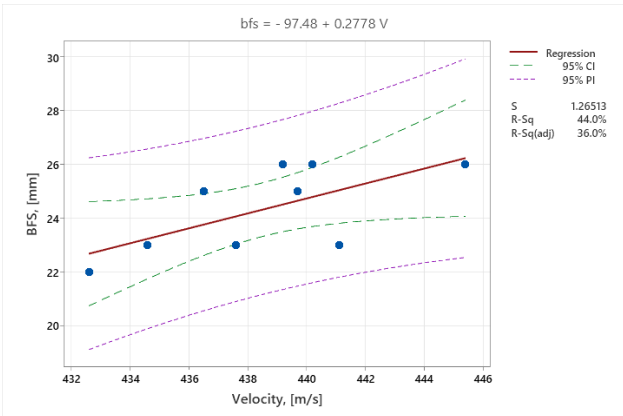


Figure 5. Fitted line plot – BFS vs. Velocity for Laboratory A

In this study, 9 mm ammunition was used with an average

impact energy of approximately 770 J. However, Level IIIA protection is also tested with .44 Magnum ammunition, which carries a nominal impact energy of 1478 J—nearly double that of the 9 mm. Consequently, BFS values could reach or slightly exceed the allowable limits, leaving it uncertain whether a failure is due to the sample's performance or the influence of the backing material's properties. While the latest NIJ 0101.07 standard addresses these issues through larger sample sizes and advanced statistical processing, it is not yet widely implemented. Many laboratories cannot yet meet these requirements for accreditation, leading manufacturers to continue opting for equipment certification under the NIJ 0101.04 version.

3.2 Ballistic test method for personal armour materials and combat clothing - STANAG 2920

The test results are presented in Table 6, where complete penetrations are denoted as CP and partial penetrations as PP. In addition to the standard method, the V_{50} value was also determined using the Probit method [5], which is known to provide reliable results, particularly in cases involving a wide range of velocity values.

Table 6. Test results for the determination of V_{50} value

Lab. code	Sample	CP, [m/s]	PP, [m/s]	V_{50}, [m/s]	V_{50}, probit [m/s]	D, [m/s]
A	1	630.3 621.6 630.8	605.7 620.9 593.4	617.1	621.3	0.7
	2	641.4 645.3 647.5	607.9 632.4 622.7	632.9	636.8	9.0
	3	639.3 643.3 645.6	608.3 621.4 619.1	629.5	630.3	17.9
B	1	619.9 608.1 617.1	608.5 601.7 603.7	609.8	608.3	-0.4*
	2	615.1 611.6 614.2	601.8 606.9 603.9	608.9	609.3	4.7
	3	613.4 613.9 614.3	608.4 609.8 613.7	612.3	613.2	-0.3*
C	1	575.6 575.5 586.5	556.7 576.8 571.6	573.8	575.1	-1.3*
	2	603.9 591.4 573.9	577.7 565.3 571.6	580.6	577.9	-3.8*
	3	569.6 572.0 589.9	603.9 583.8 566.7	581.0	580.4	-32*

In Table 6, the range between the highest partial penetration velocity and the lowest complete penetration velocity is designated as D. It is observed that negative values (marked with an asterisk) indicate an overlap between these values. This phenomenon was especially pronounced in Sample No. 3 tested at Laboratory C, where the overlap spanned almost the entire velocity range, from 569.6 m/s to 603.9 m/s. Conversely, a distinct "zone of separation" was observed in Laboratory A, where no velocity overlap occurred.

Table 7 presents the descriptive statistics for each laboratory based on both V_{50} calculation methods. The mean V50 value for each laboratory is denoted as V_{50aver}, while the standard deviation is represented by σ. The corresponding boxplot and interval plot are illustrated in Figure 6.

Table 7. Descriptive statistics for V_{50} (all values are expressed in m/s)

Lab. code	Method STANAG 2920		Probit method	
	V_{50aver}	σ	V_{50aver}	σ
A	626.5	8.32	629.5	7.78
B	610.3	1.76	610.3	2.59
C	578.5	4.05	577.8	2.65

The statistical analysis was conducted using a one-way ANOVA, with the testing location (laboratory) as the primary factor. To ensure the validity of the results, an F-test for equality of variances was first performed as a fundamental prerequisite for ANOVA. The results indicated a significant difference in variance between Laboratories A and B (p=0.043) regarding the V_{50} results calculated per STANAG 2920. However, when analyzing the Probit method results, no variances were significantly different. This highlights the robustness of the Probit method and its ability to compensate for suboptimal velocity selections. Consequently, further analysis was conducted using an ANOVA test that does not assume equal variances, supplemented by the Kruskal-Wallis non-parametric test.

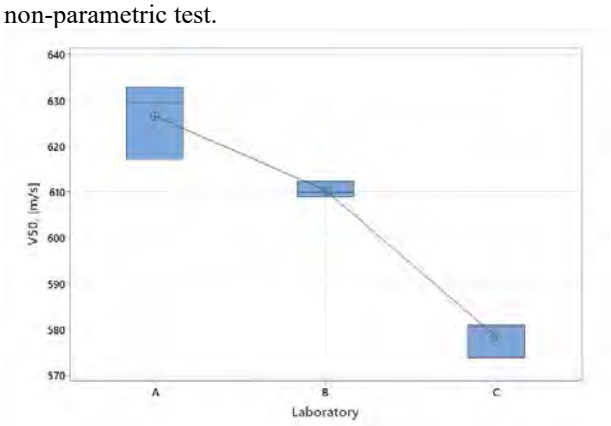


Figure 6. Boxplot and interval plot of V_{50} values

The ANOVA yielded a p-value of 0.002 (F=72.42), indicating a strong, statistically significant difference in test results between laboratories. The Games-Howell simultaneous test for differences of means confirmed that these significant differences primarily exist between Laboratory C and the other two facilities. The model's coefficient of determination R^2 was 95.29%, with a predicted R^2 of 89.40%, suggesting high precision for predicting new observations. The Ryan-Joiner test statistic for normality was 0.985 (p>0.05). Although there

appear to be large differences between Laboratories A and B, the high variance in Laboratory A combined with a relatively small sample size prevents a definitive conclusion regarding the statistical significance of the difference in their V_{50} mean values. The Kruskal-Wallis test corroborated the ANOVA results by confirming significant differences in medians between the laboratories, although this test cannot identify specific differing pairs.

These findings suggest that the "laboratory factor" plays a crucial role. This influence stems from several sources:

- Impact Velocity Determination: The velocity is not measured directly at the impact point but is calculated based on the velocity measured along the trajectory.
- Sample Mounting: The method of securing the sample and the impact of material degradation from previous hits on subsequent ballistic behavior.
- Velocity Range: The specific range of velocities used to calculate V_{50}.

The application of the Probit method resulted in values that were more consistent than those obtained through the classic arithmetic mean. To improve accuracy, laboratories should aim to achieve a sufficient number of hits to narrow the velocity range. While the STANAG 2920 prescribes a calculation based on six hits, numerical simulations suggest that a much larger number of impacts is required for a truly reliable V_{50} estimate. Furthermore, as noted in previous research [6], a higher variation in the response truth set (a larger zone of mixed results) leads to higher variation in the V_{50} estimate, reduced test efficiency, and a greater susceptibility to test-method-induced bias. This was practically demonstrated by the results from Laboratory C, particularly regarding Sample No. 3.

Due to the wide range of V_{50} values, a statistical analysis was performed in accordance with ISO 13528 to determine the measurement uncertainty. The procedure was based on the Robust Statistics Algorithm A. The results, shown in Table 8, include the standard (u) and expanded (U) measurement uncertainties. A coverage factor of k=1.96 was used for the expanded uncertainty, providing a confidence level of approximately 95%.

Table 8. Uncertainties of measurement for V50 (m/s)

Lab. code	A	B	C	Mean absolute
u	6.8	1.4	3.3	20.0
U	13.3	2.8	6.5	39.2

5. CONCLUSION

The testing conducted in accordance with the NIJ 0101.04 method indicates significant deviations in mean values depending on the specific laboratory, with the Roma Plastilina No. 1 backing material identified as the primary source of variability. This study has demonstrated and statistically proven that, for a selected threat, the correlation between projectile velocity and Backface Signature (BFS) is practically non-existent. In the given case, the linear regression and ANOVA models exhibited low R^2 values and low predictive power, likely due to the small group sizes in the study (three shots per panel). However, these models are consistent with earlier studies that also confirmed this behavior and the lack of correlation. As a result, the linear regression model cannot reliably predict BFS depth based on impact velocity. Therefore, it is recommended that laboratories, at least informatively, measure BFS for both shots (2 and 3) rather than exclusively for the shot with the higher velocity. To improve the reliability of quality assessments, manufacturers should define sample sizes and allowable BFS values based on statistically grounded test protocols, considering confidence intervals rather than isolated measurements.

The laboratory factor was also found to play a crucial role in V_{50} results, stemming from probably differences in mounting methods and velocity determination models. However, the application of the Probit method yielded results that were more consistent and robust than those obtained through the traditional arithmetic mean.

Furthermore, each laboratory must establish the measurement uncertainty of the method, which this study confirms is significantly higher than the uncertainty of the measuring instrument alone. Addressing these systemic variabilities is essential for a more accurate interpretation of ballistic protection levels, particularly when results are near the maximum allowable limits.

References

[1] BASS,C.R., SALZAR,R.S., LUCAS,S.R., DAVIS,M., DONNELLAN,L., FOLK,B., SANDERSON,E., WACLAWIK,S.: *Injury Risk in Behind Armor Blunt Thoracic Trauma*, International Journal of Occupational Safety and Ergonomics, 12(4) (2006) 429–42.

[2] NATIONAL RESEARCH COUNCIL: *Testing of body armor materials, phase III*, National Academies Press, Washington, DC, 2012.

[3] ROBERSON,C., PERCIBALLI,W., ROGERS,D., FLEMING,D.: *Consistently Depressed?*, Personal Armour Systems Symposium, Quebec, 2010.

[4] McCOY,R.L.: *Modern Exterior Ballistics: The Launch and Flight Dynamics of Symmetric Projectiles*, Schiffer Military History, Atglen, 2012.

[5] MAUCHANT,D., RICE,K.D., RILEY,M.A., LEBBER,D., SAMAROV,D., FORSTER,A.L.: *Analysis of three different regression models to estimate the ballistic performance of new and environmentally conditioned body armor*, National Institute of Standards and Technology, 2011.

[6] ANDRES,C., BOUGHERS,W.: *An analysis of V50 ballistic limit results adjusting 1st shot velocity, step-up step-down increments, truth characteristics and velocity control distributions*, Aberdeen Test Center, 2012.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

INSPECTION OF THE PROPULSION SYSTEM OF THEAUXILIARY SHIP PBR-36 "ŠABAC"

ENIS DAUTI
Serbian Army, River Flotilla Command, Novi Sad, dautienis1907@gmail.com

DRAGAN TRIFKOVIĆ
University of Defense in Belgrade, Military Academy, Belgrade, dragan.trifkovic@va.mod.gov.rs

DUŠAN CINCAR
Technical Testing Center , Belgrade, dusan.cincar@toc.urio.vs

Abstract: *The subject of this paper is the assessment of the condition of the existing propulsion system of the auxiliary ship PBR-36 "Šabac" from the aspect of the occurrence of increased vibrations on the shaft lines for power transmission to the propeller. The paper presents a detailed and comprehensive approach aimed at detecting the source of increased vibration levels in the propulsion system, which consequently led to the failure of the ship and its decommissioning, which is part of the River Flotilla of the Land Forces. By implementing control vibration measurements, using measuring instruments and diagnostic equipment, the main source of increased vibrations was determined, based on which a recommendation was given for further measures to eliminate malfunctions in the propulsion system. After examining the condition and analyzing the measurement results, the ship PBR-36 "Šabac" was sent to docking, with a focus on repairing the propulsion system. After lifting the ship to the river dock and inspecting it in a disassembled state, the assumptions obtained by the diagnostic equipment were confirmed and the source of increased vibrations in the power transmission system from the engine to the propellers was determined (defective release bearing and misalignment of the propeller shaft). After the work on replacing the defective parts and assemblies of the propulsion system, as well as bringing the eccentricity of the propeller shafts to the permissible deviation limits, the ship was launched into the water, after which control measurements were performed again. A condition assessment was performed and a comparative analysis of the deviations of the measured vibration values after docking was given in relation to the previous report and the reference values defined by the standard. Finally, a conclusion was given with a condition assessment and a proposal for further work.*

Keywords: *propulsion system, vibrations, shaft line, docking, measurement.*

1.INTRODUCTION

The auxiliary river ship PRB-36 "Šabac" is in operational use as part of the 1st River Detachment, River Flotilla, Army.

A ship of this type is designed for demagnetizing ships, forming and maintaining a command post of the river detachment rank, patrolling and controlling river routes.

Figure 1. The auxiliary river ship PRB-36 "Šabac"

The ship is powered by two four-stroke diesel engines,

both with a power of 192kW, which provide the ship with a displacement of 115.8t with a maximum speed of 22 km/h.

The ship's propulsion system consists of two parallel diesel engines, which drive one bronze propeller with fixed blades each via a coupling, cardan shaft, and propeller shaft.

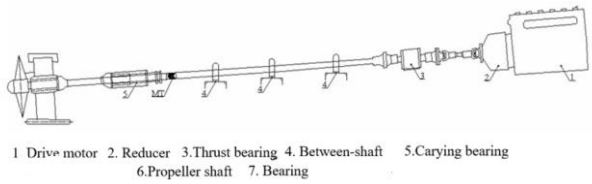


Figure 2. Sheme of propulsion

In this case, the drive motor (1), through a reducer with a coupling (2), drives a propeller with fixed blades. Between the reducer and the propeller is a thrust bearing (3), which receives the axial force from the propeller and transfers it to the ship. The propeller shaft consists of an

Corresponding author: E. Dauti
E-mail: dautienis1907@gmail.com
DOI: 10.5937/OTEH26142D

intermediate shaft supported on three roller-type support bearings (4), and a propeller shaft supported in sliding rubber bearings in the mast tube and the rudder (5).

Taking into account empirical indicators in operation, characteristic examples of failures of ship subsystems occur precisely in the system for transferring power from the engine to the propellers, which was the case here. After the malfunction of the ship's propulsion complex, as a consequence, operating parameter values outside the permissible range appeared in operation, which led to the occurrence of failures. Taking into account the experience of the crew, signs of failure in operation were identified in time, thus preventing major damage to vital ship sites.

2.FAILURE OF THE SHIP'S PROPULSION SYSTEM

During the regular operation of the ship, changes in the state of the ship's propulsion system were observed, in the form of increased vibrations transmitted to the ship's shell, a specific sound of a failed bearing, and an increased temperature of the thrust bearing housing. Based on empirical indicators of non-specific operation and characteristic deviation parameters, the ship's crew took measures to prevent further failure of vital ship systems by operating the ship within the empirically acceptable load range.

Based on the above, it was necessary to detect the degree of malfunction of the propulsion system of the auxiliary ship "Šabac", define the cause and consequences of the failure, and propose further measures that need to be taken to restore the ship to proper condition.

Detection of ship propulsion malfunctions included diagnostic measurements carried out by the Technical Testing Center, with the assistance of the ship's crew. For this purpose, and in order to obtain as realistic a picture of the problem as possible, a test run was carried out on the Danube River, in different sailing modes, which ensured a simulation of the propulsion system load in real conditions.

3.DIAGNOSTIC MEASUREMENTS

The content of the diagnostic measurements was determined by the symptomatic occurrences of failures as well as previous experience in detecting the cause of malfunctions of the ship's propulsion system on other ships of the River Flotilla. Based on the above, the diagnostic measurements included the determination of the following parameters, as a function of assessing the state of the propulsion system: torque and power measurements, linear vibration measurements and bearing temperature measurements on the ship's propeller shaft of the auxiliary ship PRB-36 "Šabac".

3.1 Measurement conditions

The measurement was carried out in upstream travel at approximately normal ship displacement (D= 110 t), on the left and right propellers. The measurement was carried out in three sailing modes, and with the restriction that the maximum mode of the propulsion engines was at 1400 rpm. The specified mode was empirically estimated as the maximum safe mode (according to the ship's commander's assessment), in which vibrations of the propulsion complex cannot cause damage to the propulsion elements and for which there is a very low risk of failure.

3.2 Measuring equipment

To measure the shaft deformation, on the basis of which the power delivered to the propeller was calculated, measuring tapes of the type XY21 6/120 (tape factor K=2.09), manufactured by Hottinger Baldwin Messtechnik, connected in a full Wheatstone bridge, were used.

To measure the number of revolutions of the propeller shafts, a tachometer TAC-10, manufactured by SPM, was used, while measurements of linear vibrations of the propulsion complex elements were performed using a portable vibrodiagnostic device Falcon and a three-axis accelerometer 0.1dB-Metravib SAS.

Measurements of the bearing housing temperature were carried out using a non-contact thermometer VOLCRAFT, IR-1000L.

4.RESULTS OF MEASUREMENT

4.1 Power measurement results on the left and right return lines

The results of the power measurements on the left and right propeller shafts were obtained based on the read values of the changes in the resistance of the tape, which is proportional to the deformation. Using the Spider 8 measurement acquisition device and the transmitted signal from the measuring tape, with prior calibration, the tangential and normal stresses were determined, and therefore the torque and transmitted power on both shaft lines.

Table 1. Result of power measurement

Nv (o/min)	Nmot (o/min)	M_v (Nm)	Pv (kW)	Nv (o/min)	Nmot (o/min)	M_v (Nm)	Pv (kW)
507	1014	668,21	35,46	509	1018	713,85	38,03
670	1240	1146,85	80,42	647	1294	1228	86,11
705	1410	1417,8	104,62	695	1390	1391,04	101,19

Nv – shaft speed, Nmot – drive motor speed, Mv – torque measured on the shaft, P v – power measured on the shaft (power delivered to the propeller).

Comparing the measured power values on the left and right shafts, it can be concluded that at lower driving modes the right engine is more loaded than the left. This is also supported by an insight into the measured values of the exhaust gas temperature of the left and right engines, where a constantly higher exhaust gas temperature of the right engine is observed compared to the left engine (the difference is up to 100 °C).

4.2 Linear vibration measurement

By analyzing the measured values given in Table 1, it is observed that the vibration level is above the values recommended by the American Bureau of Shipping (ABS) standard for vibration on marine propulsion devices, on the following elements of the return line: 1) right reducer, longitudinal and vertical, at 585 rpm; 2) left thrust bearing, longitudinal, at 605 rpm and 705 rpm; 3) right thrust bearing, vertical and longitudinal, at 500 rpm and 650 rpm.

Table 2. Result of vibration measurement

DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Reducer
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Vertical	Horizontal	Longitudinal	According to the ABS standard, the recommended permissible value is 13mm/s.
507/500	4,12/4,07	5,68/7,94	4,92/5,39	
655/650	/14,82	/6,79	/18,38	
705/690	6,13/10,15	4,79/9,11	9/8,08	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Thrust bearing
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Vertical	Horizontal	Longitudinal	According to the ABS standard, the recommended permissible value is 5mm/s
507/501	1,84/1,03	2,77/1,59	4,8/5,17	
655/651	2,1/1,69	3,99/2,29	6,25/8,2	
705/691	3,06/8,27	4,04/2,04	10,22/3,25	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Bearing
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Vertical	Horizontal	Longitudinal	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
507/502	1,24/1,19	2,11/1,76	1,85/1,3	
655/652	1,81/2,18	3,98/2,65	3,39/1,78	
705/692	4,19/3,23	5,25/3,05	4,13/4,88	
НАЗИВ УРЕЂАЈА: Бродски вратилни вод (лево/десно)				MEASUREMENT POINT LOCATION: Carying bearing
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Vertical	Horizontal	Longitudinal	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
507/503	1,95/0,91	1,24/2,08	0,73/1,18	
655/653	1,08/1,35	1,53/0,95	1,08/0,8	
705/693	3,96/2,26	1,92/1,58	1,23/1,26	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Ship's plating
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Vertical	Horizontal	Longitudinal	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
705/694	4,08	2,49	3,06	

4.3 Measuring temperature on bearings

Given the inaccessibility of the measurement site, the temperature was measured contactlessly on the bearing housing itself.

Table 3. The measured temperature values

rpm	Legft/Right thrust bearing	Legft/Right bearing between shaft	Legft/Right bearing shaft
500	33.3/42	21.3/18.1	27.3/24.4
600	34.9/43	22.6/19.5	28.7/27
700	36.6/47.2	24/19.7	34.1/29.8

4.4 Analysis of measurement results

After conducting diagnostic tests and performing measurements, and analyzing the measured operating parameters of the left and right return lines, the following conclusions were reached:

On the left shaft, values of impermissible vibration levels were observed at the thrust bearing, as well as the existence of angular misalignment in the axle assembly. The recommended permissible vibration speed value according to the ABS standard is 5 mm/s, while the measured values are up to 10.22 mm/s, at 705 rpm.

An irregularity was observed on the right shaft in the form of impermissible vibration levels on the gearbox and thrust bearing, which indicates the existence of angular misalignment of the shaft. The recommended permissible vibration speed value according to the ABS standard is 13mm/s, while the measured vibration speed values are 18.38 mm/s at 650 rpm. In addition, a malfunction of the left cardan shaft and the engine foot was detected. Taking into account the increased temperature value of the thrust bearing, the malfunction will also affect the service life, since thermal overload will occur in operating conditions over a longer period of time.

By comparing the condition values of the left and right thrust bearings, based on the frequency spectrum of axial vibration accelerations, a worse condition of the left thrust bearing is observed, which is in the zone of limited service life, i.e. there is the appearance of dominant frequencies in the zone of characteristic bearing frequencies, and the appearance of the so-called "carpet".

Taking into account that the measurements were carried out at modes lower than nominal, it is assumed that at higher modes, the values of the measured vibrations would be significantly higher.

Based on the conducted testing, docking of the ship and elimination of the malfunction were proposed.

5. DOCKING THE SHIP

Docking the ship and eliminating the malfunction, as shown in the previous chapter, was preceded by dismantling the malfunction, identifying the cause, and proposing defined works to be carried out in order to eliminate the causes and consequences of the malfunctions, identified during control measurements.

Figure 3. Docked ship

The observed defects during diagnostic testing and measurements carried out during the docking preparation process confirmed the assumptions regarding bearing defects, as well as return line misalignment.

Figure 4. Defective thrust bearing built

After troubleshooting in a disassembled state, it was determined that the thrust bearings had suffered serious mechanical damage, and that they had been exposed to operating conditions with improper cooling, and a similar condition was observed on the rolling support bearings of the return line after construction. Due to the general poor condition and worn out bearings, they were replaced with new ones. The condition of the propeller required machining, edge sharpening and balancing to bring the eccentricity within permissible limits.

Figure 5. Machined propeller bearings

As part of the final work on the docking of the ship, the painting of the ship's hull and plating was carried out.

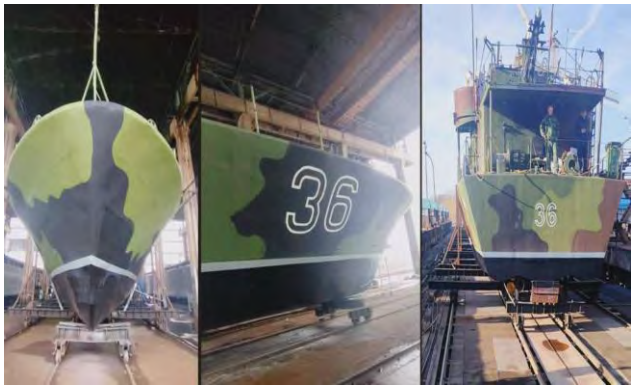


Figure 6. Appearance of the ship after docking

6. CONTROL MEASUREMENTS AFTER DOCKING

After the completion of docking and the elimination of malfunctions in the propulsion system, a test run was carried out as part of the handover works, during which control measurements were also performed to assess the condition of the ship's propulsion system.

The measurement was performed while sailing upstream at approximately normal ship displacement, on the left and right axle rows.

The measurement was performed in four sailing modes, with the note that the maximum mode was at 1590 rpm for the left main engine, and 1500 rpm for the right main engine. The specified mode was assessed as the maximum safe mode, according to the ship's commander, in which vibrations of the propulsion complex cannot cause damage to the propulsion elements.

Diagnostic measurements on the return lines included the determination of the following parameters:

- Measurement of linear vibrations on the supporting elements of the return lines;
- Measurement of temperatures on the supporting elements of the return lines.

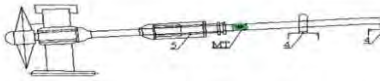


Figure 7. Position of measuring tapes on the return line

6.1 Linear vibration measurement

Linear vibration measurements were performed with the aim of establishing the vibration level, as well as comparing the measured values with the permissible values according to the American Bureau of Shipping standard.

Linear vibration measurements of the propulsion complex elements were performed using the Falcon portable vibration diagnostic device (сер.бр.10909) and the 0.1dB-Metravib SAS triaxial accelerometer, while the TAC-10 tachometer, manufactured by SPM, was used to measure the rotational speeds of the drive shafts.

Table 4. Result of vibration measurement

DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Reducer
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Вертикално	Хоризонтално	Уздужно	
495/480	4,47 / 3,06	15,99 /9,91	4,22 / 3,97	According to the ABS standard, the recommended permissible value is 13mm/s.
630/600	6,48 / 5,88	6,24 / 5,08	6,66 / 9,50	
720/75	5,95 / 9,93	5,51 / 8,47	7.29 9,37	
795/750	8,35 / 8,74	6,98 /9,2	8,35 / 8,73	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Thrust bearing
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Вертикално	Хоризонтално	Уздужно	
495/480	1,55/ 1,62	2,74 / 4,32	3,96 / 3,74	According to the ABS standard, the recommended permissible value is 5mm/s
630/600	1,92 / 1,96	4,03 / 4,01	4,00 / 3,95	
720/75	4,03 / 2,08	4,55 / 4,1	4,53 / 4,05	
795/750	3,90 / 1,97	4,95 / 3,43	7,75 / 7,69	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Carying bearing 1
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Вертикално	Хоризонтално	Уздужно	
495/480	2,55 / 1,76	5,3 1 / 4,29	5,14 / 4,03	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
630/600	4,99 / 2,59	**8,86** / 6,00	6,30 / 5,93	
720/75	5,98 / 2,49	**16,45** / 4,01	**15,35** / 6,04	
795/750	5,30 / 2,40	**10,52** / 4,30	11,10 / 7,09	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Carying bearing 2
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Вертикално	Хоризонтално	Уздужно	
495/480	1,49 / 3,14	1,64/2,38	4,70/ 5,54	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
630/600	2,03/4,29	2.94/ 2,58	5,54 /4,98	
720/75	3,95 / 5,75	6,97/ 3,63	6,89 / 7,86	
795/750	4,22 / 6,58	6,35 /4,88	6,86 / 9,56	
DEVICE NAME: Ship's shaft line (left/right)				MEASUREMENT POINT LOCATION: Carying bearing 2
rpm of the shaft (o/min)	Effective speed of vibration Vef (mm/s)			Note
	Вертикално	Хоризонтално	Уздужно	
495/480	1,93/ 1,23	3,90 / 1,86	4,49 / 1,05	According to the ABS standard, the recommended permissible value is 7 mm/s up to 13 mm/s
630/600	4.33/ 1,56	3,82 / 2,27	4,33 / 1,24	
720/75	4,77/ 2,61	6,72 / 2,93	6,97 / 2,43	
795/750	5,17 /4,37	4,95 / 5,13	7,44 / 3,63	

By analyzing the measured values given in Table 4, it is observed that the vibration level is for the most part (in terms of measurement points and modes) within the limits recommended by the American Bureau of Shipping (ABS) standard for vibrations on marine propulsion devices.

Based on the measured results, it can be concluded that the extent of the work carried out on the docking of the ship was entirely aimed at eliminating the consequences and causes of the malfunction of the ship's propulsion system.

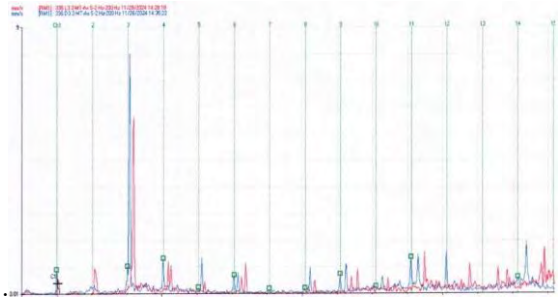


Figure 8. Frequency spectrum of the vibration on the thrust bearings

Figure 8 shows the frequency spectrum of vibration speeds, a comparison of axial vibrations on thrust bearings, for the case of the right engine 750 rpm, the left engine 795 rpm.

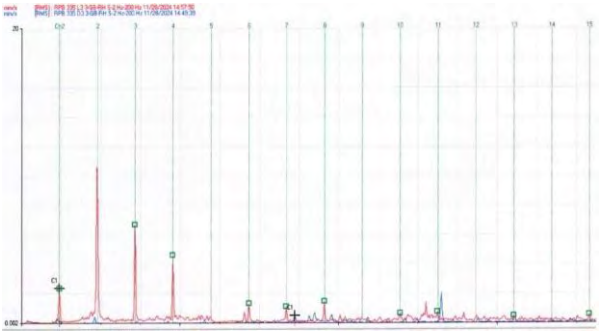


Figure 9. Frequency spectrum of the vibration on the on the first support bearing

Figure 9 shows the frequency spectrum of vibration speeds, a comparison of horizontal vibrations on the first support bearing, for the case of the right engine 675 rpm, the left engine 720 rpm.

6.2 Measuring temperature on bearings

Temperature measurement was carried out contactlessly, on the bearing housings, in four operating modes, and the VOLCRAFT IR-1000L instrument was used for this purpose.

Table 5. The measured temperature value

rpm	Left/right thrust bearning (t °C)	Left/right carying bearing 1 (t °C)	Left/right carying bearing 2 (t °C)	Left/right carying bearing 3 (t °C)
500	39,8 / 35,5	30,9 / 31,8	36,3 / 31,9	45 / 40. 1
600	37 / 31,6	29,4 ' 3 I,8	36,1 / 31,6	46,4 / 41,6
700	39,9 / 34	33,7 34,5	33,4 / 30,7	46,4 / 42.5
775	42,2 / 36,6	34,9 37,4	34,3 / 30,5	45,2 / 4 1,8

Comparing the measured temperature values, a slightly higher temperature of the last support bearings in the series is observed, but there are no results that significantly deviate from the measured temperature intervals. By reviewing the measurement results and comparing the bearing temperature values of the left and right return lines, it can be concluded that the values at different modes are approximately negligible differences for the left and right return lines, and that small differences in deviation can be attributed to the sensitivity of the measuring device.

7. EVALUATION OF MEASUREMENT RESULTS

By analyzing the measured operating parameters of the left and right shaft lines, the following conclusion is drawn:

The thrust bearing (behind the thrust bearing) of the left shaft line at driving modes above 1300 rpm had a horizontal vibration value exceeding the values recommended by the standard. The irregularity in the form of an impermissible vibration level, considering the order (second) and the direction of vibration (horizontal), indicates the probable existence of a parallel misalignment of the shaft line.

The thrust bearings on both shaft lines at maximum speed during operation had axial vibration values above the values recommended by the standard. The measured values, in accordance with previous experience and diagnostic measurements on Russian ships, are not unusual for the propulsion complex. By analyzing the frequency spectrum, it can be concluded that the higher level of vibration is caused by minor misalignments of the shaft line or unbalance of the propeller blades.

The measured bearing temperature values are within the permitted range and there are no values that significantly deviate from other measured temperatures on the return line bearings.

Empirically, from the perspective of the crew and experts from the River Flotilla who attended the test run, which lasted more than four hours at different operating modes, in upstream and downstream navigation conditions, the docking and repair work on the ship's propulsion complex met expectations and completely eliminated the consequences and causes of the failures.

It is particularly important to emphasize that early detection of failures in the operation of all subsystems is very important, because in this way the chain of failures of other subsystems and assemblies, causally and consequentially related to them, is prevented. Also, empirical understanding, understanding and "feeling" can often be crucial in cases of observing the state of health of ship subsystems, given that the assets under consideration (ships) represent complex technical systems, for which previously defined limitations and parameters do not coincide with actual operating conditions.

8. CONCLUSION

Warships are complex technical systems, which as such represent factors of a large number of subsystems.

The most important among them is the ship's propulsion system, which represents the "heart of the ship".

Ensuring the health of a ship's propulsion system involves a complex technical process of observing a number of parameters, as well as monitoring the health of components and return line elements. For this purpose, diagnostic tests are often carried out, in which measurements are carried out to determine quantitative indicators and assess the condition of the tested system, which was the case here.

The role of the crew was very important in the early detection of the failure of ship systems, so thanks to the timely reaction and diagnostic measurements, the malfunction of the ship's propulsion system of the auxiliary river vessel "Šabac" was confirmed, while maintaining the functional integrity of the ship by applying a limited mode of use (load).

After determining the malfunction of the ship's propulsion system of the auxiliary river vessel "Šabac" and analyzing the results of diagnostic measurements, the ship was docked and comprehensive work was carried out to eliminate the causes and consequences of the failure.

During the docking process, the ship's propulsion system components were disassembled and inspected, which allowed for a clear understanding of the causes and consequences of the malfunction. After the malfunction was identified, all defective components and assemblies of the ship's propulsion system were replaced, the return lines were processed to bring the eccentricity within the limits of permissible deviations, and the propellers were balanced.

After eliminating the observed malfunctions in the ship's propulsion system and the works envisaged by the scope of docking, a test run was carried out as part of the handover works, during which control measurements were carried out on the ship's propulsion system.

By analyzing and comparing the results of control measurements before and after docking, we conclude that the values of the measurement parameters after docking are within the limits of permissible deviations. Also, the empirical observation of the state of the ship's propulsion system by the ship's crew fully confirmed the drastic difference in ensuring the ship's maneuvering and propulsion characteristics after docking, which is in the function of ensuring the expected values.

References

[1] TEHNICAL TESTING CENTER: *Report number 12-3763, April 2024.* Belgrade.

[2] TEHNICAL TESTING CENTER: Report number 12-3769, December 2024. Belgrade.

[3] IVANOVIC,G., STANIVUKOVIC,D., BEKER,I., *Reliability of tehnical systems*, Facultz of Mechanical Engineering in Belgrade.

[4] ADAMOVIC, Z., *Total maintenance of tehnical systems*, Tehnical Faculty „Mihajlo Pupin“ Zrenjanin.

[5] TRIFKOVIC, D., NIKOLIC, R., PETROVIC, Ž., *Measurement of torsional vibrations by using strain gages,* Military technical gazette 3–4/2005.

[6] Chetan K Changale, Prof (Dr.) Anil R Sahu, *Vibration Analysis of a Rotating Shaft*, International Journal for Research Trends and Innovation, Volume 10, Issue 5 May 2025.

[7] Technical documentacion of the ship.

[8] HBM Spider8 Operating Manual.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

MECHANICAL QUALIFICATION OF SINGLE SOURCE COMMERCIAL OFF THE SHELF (COTS) ASSEMBLIES FOR AEROSPACE AND DEFENCE ENVIRONMENTS

KONSTANTIN KOMATINA
Vlatacom Institute, Belgrade, konstantin.komatina@vlatacom.com

Abstract: *This paper establishes a test-based validation protocol to certify the "fitness for use" of a commercially available product that is subjected to aerospace and defense grade mechanical operating loads.*
Methodology rests on three pillars: Load characterization according to relevant standards; Testing based on those standards and Post load functional Verification.

Keywords: *mechanical qualification, COTS mechanical validation, structural integrity defence and aerospace MIL-STD*

1. INTRODUCTION

This paper establishes a rigorous validation protocol to certify the "fitness for use" of a specific commercial assembly subjected exclusively to aerospace-grade mechanical operating loads. In scenarios where other stressors like thermal and radiation are negligible, the primary risk to commercial hardware is structural failure or functional degradation due to launch-induced vibration and/or high-magnitude shock loads.

Commercial hardware is often chosen as a downgrade replacement for aerospace and defense grade components when such components are not readily available to design bureaus for various reasons. In aforementioned circumstances design bureaus must either rely that commercial components are sufficient for serious applications, or design and implement a validation protocol that will subject COTS to operating loads or verify failure modes in order to ensure reliable operation when time comes.

2. OPERATIONAL PROFILE AND WORK SCOPE

The test specimen in question is chosen due to being readily available, modestly priced and technologically familiar to engineers and technicians of various levels of expertise. This specific connector is available for purchase from web stores (1) for example.

Operational requirements of mentioned component are relatively unusual in practice and require clean separation and extraction during very short time period.
For this specific connector there are no available quality certificates, and certainly there are no datasheets that address the rigorous operating loads

Figure 1 – Push pull Connector

Thus, the next steps can be summed up as such:

- Define regulatory framework
- Define experimental design
- Define acceptance criteria

And in conclusion provide clear boundaries in COTS validation and applicability of industry standards to a specific engineering problem under severe test conditions.

Conclusion of this paper will be in comments about reliability of selected product according to MIL-DTL

Corresponding author: K.Komatina;
E-mail: konstantin.komatina@vlatacom.com

DOI: 10.5937/OTEH26143K

3. REGULATORY FRAMEWORK

Starting point for any validation protocol is to define the regulatory framework that covers testing of the components. Unfortunately, what is readily available are standards that cover a wide range of tests in factory acceptance phase of the product development, such standards are MIL-STD-202 (2) for electronic and electrical components and parts, and MIL-STD 810 (3) for an even wider group of products.

These standards are not applicable to the scenario that is described.

What can be used is Detail specification MIL-DTL-24308G. This specification covers requirements for nonenvironmental, polarized shell, miniature, rack and panel connectors having pin and socket, for a -55 ^{0}C to +125 ^{0}C range (2). This range is compliant to operating temperature range per MIL-STD-810 (3).

4. EXPERIMENTAL DESIGN, ACCEPTANCE CRITERIA

Purpose of using aforementioned standard is to define a range of mating and unmating forces that are to be expected. These forces must be replicated during the test and based on those values, a simple PASS/FAIL Criteria is defined.

Mating forces are dependent on shell size of the connector and also its internal design (2).

There are alternative ways to verify (4) connector mating forces, however, such methods are more suited to the designers of the specific connector and not to end users or design enginners that intend to incorporate such products in their design.

Of equal importance is to define unmating speeds. Unmating speed in operating conditions cannot be replicated in laboratory conditions, and in this particular case would require a moderately expensive rig and a lot of resources to perform. MIL standard offers us a range of unmating speeds that can be replicated.
Those are given in a table bellow

Table 1 - Table from MIL-DTL

	low end	high end	Note
Unmating Force [N]	3	27	Per MIL DTL
Unmating speed [mm/s]	0.4	4	Test performed at unmating speeds of 2mm/s

During the experiment following items are used:

- Pair of calibrated 50gr, 100gr and 500gr weights for load cell calibration
- Number 6 Allen key
- Number 17 wrench
- Commercially available tensile test rig that has a load cell that measures forces up to 1000N

Figure 2 - Calibrated weights

Figure 3 - Number 6 Allen key

Figure 4 - Number 17 wrench

Figure 5 - Tensile test machine with load cell

Acceptance criteria is based upon visual confirmation that the push pull connector has extracted under all angles of attack. During the test, connector is subjected to loads under angles of attack of 0, 15, 30 and 45 degrees to verify across the entire expected range of operation. In order to achieve such alignment, a number of fixtures are manufactured.

Inside the connector the pins and soldered joints should not be damaged.

Damage to the outer shell is allowed.

Figure 6 - 45 degrees of extraction

Figure 7 - Connector pins

5. CALIBRATION, RESULTS

Before the start of the test load cell calibration is needed to produce a calibration curve. This curve will be used later on the determine correct force that was applied during the test

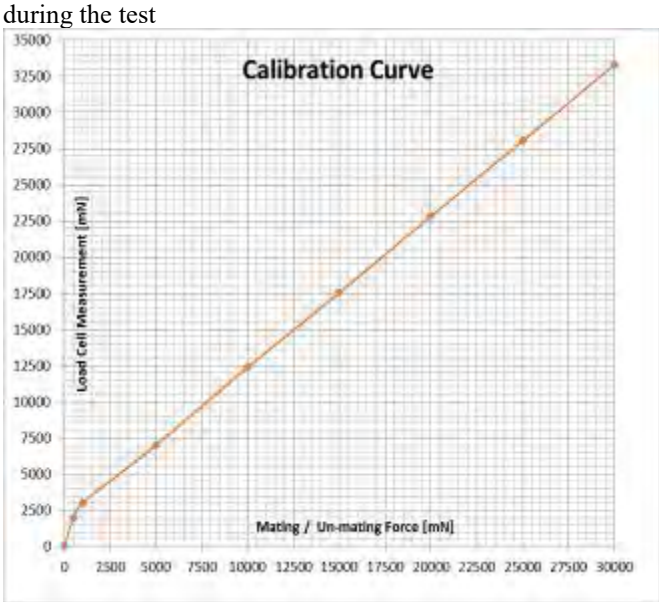


Figure 8 - Calibration curve

Force that is acting upon the connector is divided in 2 components – Fy component is the mating force and Fx component is the lateral component that will come into play as the angle of attack rises.

$$Fx = F * \sin(a) \quad (1)$$

$$Fy = F * \cos(a) \quad (2)$$

Only three of four mentioned angle of attack results will be shown in this chapter. Case number 4 will be discussed in the Conclusions section

Table 2 - Case 1 AOA 0 degrees

Drag angle [deg]		0	
#	Load cell	Unmating Force	Pass/Fail
/	N	N	/
1	14	14	Pass
2	14	14	Pass
3	15	15	Pass
4	17	17	Pass
5	17	17	Pass
6	18	18	Pass
7	18	18	Pass
8	18	18	Pass
9	19	19	Pass
10	20	20	Pass

Table 3 - Case 2 AOA 15 degrees

Drag angle [deg]		15		
#	Load cell	Unmating Force	Lateral Force	Pass/Fail
/	N	N	N	/
1	18	17	5	Pass
2	22	21	6	Pass
3	22	21	6	Pass
4	22	21	6	Pass
5	24	23	6	Pass
6	25	24	6	Pass
7	21	20	5	Pass
8	21	20	5	Pass
9	23	22	6	Pass
10	22	21	6	Pass

Table 4 - Case 3 AOA 30 degrees

Drag angle [deg]			30	
#	Load cell	Unmating Force	Lateral Force	Result
/	N	N	N	/
1	793	687	397	Pass
2	491	425	246	Pass
3	491	425	246	Pass
4	393	340	197	Pass
5	345	299	173	Pass
6	351	304	176	Pass
7	322	279	161	Pass
8	302	262	151	Pass
9	311	269	156	Pass
10	267	231	134	Pass

6. ANALYSIS AND CONCLUSION

Tabelar results are provided for only 3 of 4 tested cases.

- Case # 1 verified values retrieved from MIL-DTL (2), however prescribed durability of 500 cycles of mating and unmating wasn't tested as being deemed incompatible with real world operation
- Case #2 provided an insight in safety margins that were used when designing the Detail specification given that with the added increase in force, upper limit from MIL-DTL wasn't reached

From both these cases it can be concluded that MIL-DTL is very useful when verifying operation up to 15 degree pull angles.

Other two cases, Case #3 and Case #4 go beyond the industry standard, and are considered separately. Comments about the failure modes and effects on reliability will be provided in section below.

Case #3

Case #3 shows a great increase in force that results in damage to the lanyard that is used to pull the connector – see the figure bellow, and to the leaf springs that give the connector the push pull capability

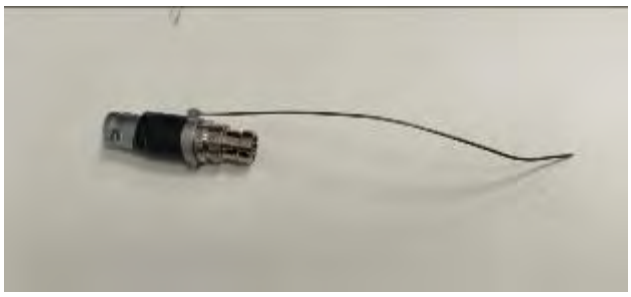


Figure 9 - Lanyard failure

Failure of the lanyard or damage to the leaf springs is not covered by MIL-DTL. Main failure mode identified by Detail Specification is damage to the pins and soldered connections inside the connector – **In this case, they are intact**

Reliability Inference

Connector can be still used with minor modifications to the lanyard – thicker cable and swaged joint will prevent the failure of the lanyard.

Thus connector should still be capable of MIL-DTL recommendation of 500 mating and unmating cycles.

However this would need to be verified.

Case #4 – 45 degree angle

This load case is the most severe of all provided.

To prepare the sample for the test, new lanyard must be engineered.

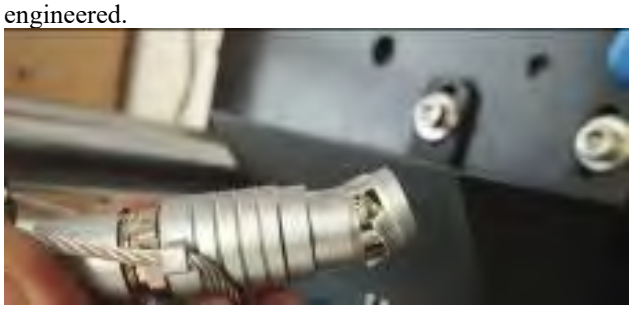


Figure 10 – Case #4 – Connector break

Table 5 - Case #4 AOA 45 degrees

Drag angle [deg]		45		
#	Load cell	Unmating Force	Lateral Force	Result
/	N	N	N	/
1	728	515	515	Fail
2	907	641	641	Fail

Distinct failure mode can be observed, lower part of the shell has completely broken, however the pins and soldered connections inside are still usable – unfortunately this connector is not.

Due to severe damage further tests at 45 degrees pull angle were not performed.

Reliability Inference

This connector is not usable and must be immediately replaced with one in working condition.

In this case, 500 mating cycles that are recommended in MIL-DTL are not achievable and this connector is single use only.

References

[1] Push pull connectors. Shenzhen : Shenzhen Newourconn Technology Co Ltd., 2026.

[2] Department of Defense, USA. Detail Specification, Connectors, Electric, Rectangular, Nonenvironmental, Miniature, Polarized Shell, Rack and Panel, General specification. Washington DC : DOD, 2009.

[3] Department of Defence, USA. U.S Department of War Test Method Standard, Environmental Engineering Considerations and Laboratory Tests. Washington DC : Department of Defence USA. 810-G.

[4] SamTec. Calculating Connector Mating Forces. s.l. : SamTec, 2018. Blog post.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMBINED APPLICATION OF FTA AND FMEA METHODS FOR FAILURE ANALYSIS OF TECHNICAL SYSTEMS

DOBRIVOJE ĆATIĆ
Faculty of Engineering, University of Kragujevac, 34000 Kragujevac, 6 Sestre Janjić Street, Serbia, caticd@kg.ac.rs

VLADIMIR ĆATIĆ
HTEC Group, Bulevar Milutina Milankovića 7đ, Belgrade 11070, Serbia, vladimir.catic@htecgroup.com

Abstract: *This paper presents the results of research related to the combined application of the Fault Tree Analysis (FTA) and Failure Modes and Effects Analysis (FMEA) methods for analyzing the reliability and safety of technical systems. FTA is a deductive (top-down) method in which, for a defined top event, the causal events that lead to it are identified. FMEA is an inductive (bottom-up) method for conducting reliability and safety analysis from lower to higher system levels. The individual application of these methods is time-consuming and does not provide a complete insight for reliability analysis and risk assessment. The combined application of FTA and FMEA enables the full utilization of the advantages of both methods while minimizing their limitations. Through the sequential application of FTA and FMEA across structural levels and their recursive application from the top to the bottom of the technical system structure, critical components that limit reliable and safe operation can be rapidly identified. The combined application of FTA and FMEA is illustrated through a case study involving the identification of critical components in the control system of light commercial vehicles. The paper concludes with directions for future research in this field.*

Keywords: *failure analysis, FTA, FMEA, combined application.*

1. INTRODUCTION

Reliability analysis of technical systems is based on the failure analysis of their components. Numerous methods have been developed to analyse failures in technical systems and their components. Among these methods, Fault Tree Analysis (FTA) and Failure Modes, Effects, and Criticality Analysis (FMECA) are the most widely used. [1, 2].

Fault Tree Analysis is a deductive analytical method in which, for a predefined top event representing the failure of the considered structural unit or system, the causal events leading to that failure are identified. The basis of Fault Tree Analysis is the representation of physical systems by means of structural logical diagrams.

The FTA method was developed in the early 1960s in the United States for the reliability and safety assessment of military systems [3, 4]. The method was conceived by H. A. Watson of Bell Telephone Laboratories**.** During 1961 and 1962, Watson developed and applied the method to the safety analysis of missile launch systems for the U.S. Air Force. Since the mid-1960s, FTA has been extensively applied to reliability and safety assessment, as well as fault diagnosis, of a wide range of complex technical systems. The method is particularly suitable for analyzing systems whose failures could result in catastrophic effects for human life and the environment.

Failure Modes and Effects Analysis (FMEA) is an inductive method for performing a qualitative analysis of the reliability and safety of a system, proceeding from the lowest to the highest system level [5]. It is based on the examination of all potential failure modes of the system components and their effects on the overall system, with the objective of identifying those failure modes that have the most severe impact on system performance.

FMEA was originally developed for military applications in the United States as a reliability assessment technique for determining the consequences of different failure modes in technical systems. As an official document in the form of a U.S. military standard, the method dates back to November 9, 1949, under the designation MIL-P-1629, entitled Procedures for Performing a Failure Mode, Effects and Criticality Analysis.

The standalone application of Fault Tree Analysis (FTA) and Failure Mode and Effects Analysis (FMEA) methods for risk assessment is time-consuming and does not provide a comprehensive insight into the criticality of system components and their failure modes. Combining these methods leverages their respective strengths while simultaneously minimizing their limitations.

A significant contribution to this field is presented in [6], which proposes a method where FTA and FMEA are applied recursively across different system levels. In this approach, FTA is first utilized to identify potential causes of system failures, after which FMEA is applied to assess the criticality of the identified failures. This procedure is repeated at various levels of the system hierarchy, thereby enabling a detailed analysis of system behavior and a more efficient identification of critical components.

Corresponding author: D. Ćatić;
E-mail: caticd@kg.ac.rs
DOI: 10.5937/OTEH26144C

Further development of integrated methods is presented in [7], which proposes an integrated FTA-FMEA model for risk analysis in complex engineering systems. In this approach, the minimal cut sets identified by FTA are used to determine the importance of individual events within the system. The obtained values are then used to modify the Risk Priority Number (RPN) index within the FMEA, resulting in a more accurate risk assessment and better identification of critical system components.

The combined application of FTA and FMEA methods in this paper is illustrated through a criticality analysis of steering system components in light commercial vehicles. Through the sequential and recursive application of FTA and FMEA methods across the structural levels of the system, the components that restrict safe and reliable operation were determined.

2. FAULT TREE ANALYSIS

An analysis of relevant FTA standards and the literature presented in [4] resulted in the following FTA methodology:

1. Define the system of interest and specify any assumptions to be used in the analysis.
2. Establish the boundaries and objectives of the system,
3. Define the top event of the fault tree. Depending on the analyzed system, it can be general, in the form of a system failure, or a specific case involving only a few failures of the system or its components.
4. Systematically collect and analyze system data (including documentation, calculations, catalogues, service instructions, user complaints etc.).
5. Verify and refine the fault tree. If the fault tree does not accurately reflect the actual system, if all significant events are not included, or if there is no logical relationship between the basic events and the top event, collect additional system data and revise the fault tree accordingly.
6. Perform a qualitative and/or quantitative analysis. Depending on the objective of the analysis, perform a qualitative and/or quantitative evaluation once the fault tree has been verified. For quantitative fault tree analysis, determine the minimal cut sets, which form the basis for most quantitative analyses.
7. Determine appropriate corrective measures. Based on obtained and adopted results through the fault tree, one should provide suggestions for corrective measures aimed at eliminating the perceived defects or proposals for alternative solutions.
8. Whenever the project changes, update and verify the fault tree and the associated analysis.
9. Document the analysis process and follow up on the implementation of identified corrective actions.

In the following basic approach, the development tree, analysis and eventual recommendations for corrective actions are presented as separate steps. In actual practice, there is a great deal of interactions between the steps listed. As a result, the fault tree that evolves includes additions and/or changes reflecting an improved understanding of the various faults.

A fault tree is constructed using event symbols, logic gates, and transfer symbols [5]. A variety of event symbols are employed to distinguish between intermediate (developed) events and basic initiating events. Intermediate events are represented by rectangles. Among the symbols used for basic events, the most common are the circle, representing a component state determined by its inherent characteristics, and the diamond, representing an undeveloped event. The logic gate symbols in a fault tree define the logical relationships and dependencies between events at different hierarchical levels. For example, an OR gate produces an output if one or more input events occur, whereas an AND gate produces an output only if all input events occur. Transfer symbols facilitate the construction of complex fault trees by linking the main fault tree to one or more subtrees.

3. FAILURE MODES AND EFFECTS ANALYSIS

FMEA is fundamentally a procedure for the qualitative assessment of the reliability of technical systems. A logical extension of FMEA is the quantification of failure-related parameters of system components and the evaluation of their criticality. This extension, which focuses on assessing the degree of criticality of individual components with respect to the system or its mission, is referred to as Criticality Analysis (CA). Criticality is generally defined as a relative measure that combines the severity of the consequences of a failure mode with its frequency of occurrence. Combining FMEA and CA results in Failure Modes, Effects, and Criticality Analysis (FMECA).

FMECA can be applied at various stages of a product life cycle. According to the literature, it is most commonly used to analyze technical system components during the product development phase and to evaluate operational data collected during product service.

When failure data for the components of a technical system are available from operational experience, or when the relevant failure indicators can be reliably estimated, the quantitative FMECA procedure is performed in three steps [1, 8]:

1. Determination of criticality of failure mode j of component i is to be done by categories of failure effects k (k = 1, 2, 3, 4), using:

$$C_{ij}^{(k)} = \alpha_{ij} \cdot \beta_{ij}^{(k)} \cdot \lambda_i \cdot t_i \tag{1}$$

where α_{ij} is a relative portion (frequency measure) of failure mode j of component i ($0 \le \alpha_{ij} \le 1$, $\sum_j \alpha_{ij} = 1$); $\beta_{ij}^{(k)}$ is conditional probability that failure mode j of component i will cause category k failure effect according to the adopted classification (values are taken from Table 1; $\lambda_i = \sum_j \lambda_{ij}$ is failure rate of component i; λ_{ij} is part failure rate of mode j of component i, and t_i is operating time of component i.

Table 1. Values of conditional probabilities

Degree of occurrence of the *k*-th failure effect category	$\beta_{ij}^{(k)}$ [-]
Certain event	1
Probable event	0.1 ... 1
Most probably would not happen	0 ... 0.1
Practically doesn't happen	0

Calculated values of $C_{ij}^{(k)}$ are the initial basis for determination of other quantitative properties of component criticality. They make it possible to rank the component failure modes according to effects in order to evaluate the most critical system failure modes from the aspect of safety.

2. Determination of failure criticality of component *i*, which causes the *k*-th category of failure effects:

$$C_i^{(k)} = \sum_j C_{ij}^{(k)} \tag{2}$$

Calculation of $C_i^{(k)}$ enables the identification of the most critical components whose failures lead to specific categories of effects.

3. Criticality of the *k*-th system effect category, determined by summing the criticalities of all component failure modes for the specified effect category:

$$C_k = \sum_i \sum_j C_{ij}^{(k)} \tag{3}$$

Calculated values of C_k are statistical indicators of the representation rating of the individual category of effects.

For quantitative FMECA of technical systems, a computer program was developed as part of this study.

This software consists of a routine for calculation of failure modes criticality and effect categories and their ranking, a subroutine for transformation of components of arbitrary array into descending series and a routine for forming of input files.

4. COMBINED APPLICATION OF FTA AND FMEA METHODS

The standalone application of FTA and FMEA methods has certain limitations.

The way the top event in the fault tree is defined greatly influences the success of the analysis. If the event is too general, the analysis becomes cumbersome and unmanageable. Conversely, if the event is too specific, the analysis does not provide sufficient insight into the system. Constructing a fault tree for complex technical systems is time-consuming, leading to delayed results during the product development phase. The focus of fault tree analysis is placed on analyzing the root causes of failures that lead to system failure. The consequences of failures on the potential system breakdown, people, and the environment are not considered. In addition, the FTA method has limited capabilities regarding component criticality analysis.

FMEA focuses on the analysis of individual components and often neglects complex interdependencies between various events within the system. FMEA is not suitable for application in complex technical systems.

Consequently, numerous researchers have proposed integrated approaches that combine FTA and FMEA to leverage the advantages of both methods and mitigate their individual limitations.

The combined FTA and FMEA method utilizes deductive and inductive approaches as needed. The constructed fault tree for the adopted top event, the constituent components, and their functions are considered in parallel. The objective of the method is to perform a comprehensive analysis of the causes, modes, and consequences of component failures based on the constructed fault tree and the completed FMEA worksheet. The fault tree enables a complete analysis of cause-and-effect relationships between events. Procedures defined within the FMEA method are used to analyze the criticality of component failure modes. It should be emphasized that these procedures significantly differ between FMEA based on product operational data and FMEA in the development phase. In the product development phase FMEA, the following parameters are determined for each identified component failure mode: Severity, Occurrence, and Detection. Based on these parameters, the Risk Priority Number (RPN) is calculated using the formula:

$$\text{RPN} = \text{Severity} \times \text{Occurrence} \times \text{Detection} \tag{4}$$

The obtained RPN values are used to rank the identified component failure modes and to prioritize the implementation of corrective actions.

The results of the quantitative fault tree analysis, in the form of minimal cut sets of events, enable the correction of RPN values and an improved risk assessment.

The combined method is particularly suitable for analyzing component failures in complex technical systems. It enables the precise identification of critical components. The proposed integrated approach allows for a more efficient system reliability analysis by combining the benefits of both approaches.

The columns of the FMEA worksheet regarding failure modes, causes, and consequences of component failures are filled out based on the constructed fault tree. This implies that the results of the FTA method are used as an input for the FMEA analysis, representing a sequential approach. To reduce the time required to obtain results on the criticality of complex technical system components and to decrease analysis complexity, it is expedient to conduct the described analysis procedure recursively across the structural levels of the system.

In the combined application of FTA and FMEA methods, the synergistic effect is fully realized. Namely, the contribution of the integrated application of FTA and FMEA in terms of work efficiency and effectiveness, as well as minimizing the time required to achieve the desired results, is significantly greater than the sum of the results of their individual applications.

5. CRITICALITY ANALYSIS OF STEERING SYSTEM COMPONENTS USING THE COMBINED FTA AND FMEA METHODS

5.1. Criticality analysis of system components

As part of the master's thesis [9], the research findings of which were published in [10, 11, 12, 13], the failures of the steering system of light commercial vehicles were analyzed using the FTA and FMEA methods. By constructing the fault tree, a cause-and-effect relationship was established between system failures and the failures of components at the lowest structural level. This approach made it possible not only to identify the most common failure modes of the system components but also to determine their ultimate effects.

With the objective of identifying the critical components that limit the reliable and safe operation of the steering system, a quantitative FMECA of the steering system components of commercial vehicles was performed using operational data. This analysis was based on a comprehensive dataset, part of which is presented in Table 2. The third column of Table 2, listing the failure modes of the system components, was derived from the fault tree. Furthermore, the effect categories assigned to the individual component failure modes (Column 7) were determined based on the intermediate events identified in the fault tree. The procedure used to collect the remaining data presented in Table 2 is described in detail in the cited references.

Table 2. Excerpt from the data table for the quantitative FMECA of the steering system of light commercial vehicles

Component name	Q [-]	Failure mode	Failure mode code	Rel. rate α_{ij} [-]	Loss prob. β_{ij} [-]	Final effect	t_{srj} [km] x10^3	t_{sri} [km] x10^3
Steering wheel 61001	1	Damage Insufficient tightening of screws	N.06 N.31	0.02 0.01	1.0 0.1	*k*.4 *k*.2	20 3	873.4
Steering shaft bushing 61005	1	Cracked bushes Increased clearance	N.06 N.35	0.3 0.7	1.0 1.0	*k*.3 *k*.3	200 100	130
Rubber insert 61011	1	Cracking	N.06	1	1.0	*k*.3	200	200
Worm shaft 61103	1	Damage of the teeth Cracking - fracture Damage of grooves	N.05 N.06 N.07	0.006 0.0005 0.001	1.0 1.0 0.8	*k*.3 *k*.1 *k*.3	300 350 200	895.4
The lever of the longitudinal tie-rod with ball joint 62110	1	The deformation of the lever Cracking - fracture Damage of the thread Loose connections Increased clearance	N.02 N.06 N.07 N.31 N.35	0.0005 0.0007 0.003 0.001 0.9948	0.6 1.0 0.8 1.0 1.0	*k*.3 *k*.1 *k*.2 *k*.2 *k*.3	30 100 100 15 100	99.9
Tie-rod ball joint 62210	1 1 1	Body joint deformation Fracture of the eyeball or the ball joint body Damage of the thread Loose connections Increased clearance	N.02 N.06 N.07 N.31 N.35	0.0002 0.0007 0.003 0.001 0.9951	0.3 1.0 0.8 1.0 1.0	*k*.3 *k*.1 *k*.2 *k*.2 *k*.3	40 100 100 15 100	99.9

By processing the data presented in Table 2, using the software program for conducting quantitative FMECA, the following results were obtained:

- degree of criticality of the failure modes of the steering system components, regardless of their effects,
- degree of criticality of the steering system components according to the effects of their failure modes, and
- degree of criticality of the final effects of failures of the steering system components.

Based on the degree of criticality of the final effects of failures in the steering system components, 99.38% of the components are associated with failure modes classified within the third effect severity category (the input and output parameters of the steering system exceed the permissible limits; examples of such deviations include increased backlash, steering wheel torque requirements that exceed normal operating values and misaligned wheel geometry resulting from the deformation of steering linkage components). Table 3 presents an excerpt from the criticality ranking of the steering system components according to failure effects, specifically referring to the components assigned to the third effect severity category.

Table 3. Degree of component criticality by effects on *k*.3

No.	Code	Component name	Criticality
1	62210	Tie-rod ball joint	0.5995E-03
2	62110	Lever with ball joint	0.5995E-03
3	61005	Steering shaft bushing	0.4608E-03
4	61011	Rubber insert	0.3003E-03
5	61110	Shaft with segment	0.3986E-05
6	61002	Steering shaft	0.8208E-06
7	61107	Slide bearing of output shaft	0.6707E-06
8	61104	Roller bearing of the steering head	0.5429E-06
9	61012	Flange on the worm shaft	0.4567E-06
10	61103	Worm shaft	0.4558E-06
11	61004	Steering shaft bracket	0.4027E-06

Figure 1 shows Pareto analysis of criticality of the steering system components with failure modes that have a third category of effects.

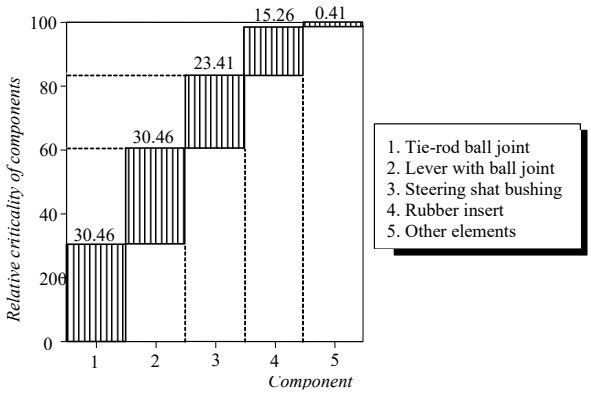


Figure 1. Pareto analysis of criticallity rate of components for category of effects *k*.3

Based on the results of the program for a quantitative analysis of the degree of criticality and Figure 1, it can be concluded that the most critical components of steering system in terms of durability and safety are: tie-rod ball joint, the lever of the longitudinal tie-rod with ball joint, steering shaft bushing and rubber insert of coupling. All other components are within 0.41% of criticality.

5.2. Criticality analysis of the tie-rod ball joint

Based on the previous analysis, it was determined that the tie-rod ball joint and the lever with the ball joint are the most critical components of the steering system. It should be emphasized that the criticality of the lever with the joint is conditioned by the criticality of the joint. In the further study, by applying a combined FTA and FMECA method, the criticality of failure modes and parts of the tie-rod ball joint was determined [9, 13]. To identify potential failure modes of the tie-rod ball joint, a fault tree was formed, shown in Figure 3. In this way, potential failure modes were recorded and their effects determined. The results of the FTA analysis were used to complete Table 4 with data for quantitative FMECA of the tie-rod ball joint parts.

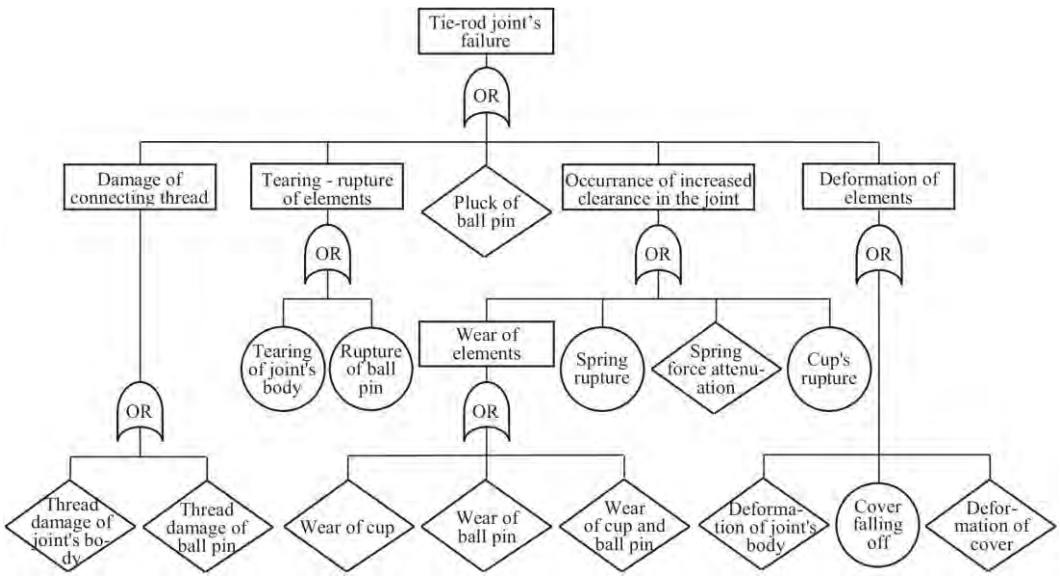


Figure 2. Independent failure tree of the light commercial vehicle steering tie-rod ball joint

Table 4. Basics of the FMECA procedure for components of the tie-rod ball joint for light trucks

Component name	Q [-]	Failure mode	Failure mode code	Rel. rate α_{ij} [-]	Loss prob. β_{ij} [-]	Final conseq.	t_{srj} [km] x10^3	t_{sri} [km] x10^3
Joint body 62211	1	Deformation Tearing of joint body Thread damage	N.02 N.06 N.07	0.0001 0.0007 0.003	0.3 1.0 0.8	*k*.3 *k*.1 *k*.2	200 200 100	≈100
Ball pin 62212	1	Rupture (tearing)of ball pin Thread damage Pluck of ball pin Wear	N.06 N.07 N.29 N.77	0,0007 0.0030 0.0002 0.9961	1.0 0.9 1.0 1.0	*k*.1 *k*.2 *k*.1 *k*.3	200 100 100 100	≈100
Cup 62213	1	Cup rupture Wear	N.06 N.77	0.003 0.997	0.7 1.0	*k*.3 *k*.3	150 100	≈100
Spring 62214	1	Spring rupture Spring force attenuation	N.06 N.12	0.0008 0.001	1.0 0.8	*k*.3 *k*.3	200 20	100
Cover 62215	1	Deformation Cover falling off	N.02 N.29	0.0002 0.0004	0.7 1.0	*k*.3 *k*.3	100 200	≈100
Sealing cap 62216	1	Total rupture Falling off Appearance of small rifts	N.06 N.29 N.37	0.002 0.007 0.004	1.0 0.9 0.2	*k*.4 *k*.4 *k*.4	30 150 20	99.89

By processing the data presented in Table 4 using the software program for conducting quantitative FMECA, the following results were obtained:

- degree of criticality of the tie-rod ball joint components failure modes, regardless of their effects,
- degree of criticality of the tie-rod ball joint components according to the effects of their failure modes, and
- degree of criticality of the final effects of failures of the tie-rod ball joint components.

Based on the degree of criticality of the ultimate failure effects of the tie-rod ball joint component, 99.21% of the total criticality sum of the components refers to failure modes of components with the third category of effects. Table 5 presents the ranked criticality values of parts for effect category *k*3.

Table 5. Degree of component criticality by effects on *k*.3

No.	Code	Component name	Criticality
1	62213	Cup	0.5986E-03
2	62212	Ball pin	0.5965E-03
3	62214	Spring	0.9598E-06
4	62215	Cover	0.3239E-06
5	62211	Joint body	0.1796E-07

Figure 3 shows distribution of criticality degree of the tie-rod components for effect category *k*.3. As it may be seen in Figure 3, failure of the tie-rod ball joint in the largest number of cases appears due to failure of the cup or the ball pin, while only 0.11% is due to other components.

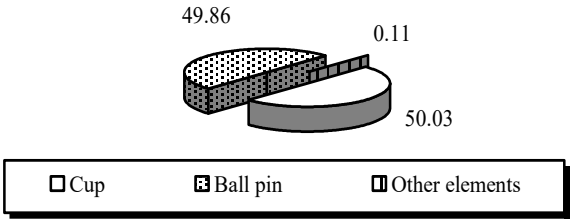


Figure 3. Distribution of degree of criticality for components of the steering system tie-rod ball joint for category of effects *k*.3

Analysis of the causes of failures in steering system tie-rod ball joint components can help identify appropriate measures to minimize or eliminate the causes of failures, or to reduce their effects.

6. CONCLUSION

The research results show that the integrated application of FTA and FMEA methods represents an effective approach for reliability analysis and risk management in complex technical systems. The combination of FTA and FMEA enables the identification of critical system components as well as more precise risk ranking based on the system structure and the characteristics of individual failures. The proposed methodology can be applied in various industrial fields, including manufacturing systems, electronic devices, and complex technical systems. As a direction for future research, the integration of this methodology with other risk analysis methods, such as probabilistic models, fuzzy logic, and Bayesian networks, is proposed to enable even more accurate quantitative risk assessment in complex technical systems.

This research was supported by the Faculty of engineering, University of Kragujevac [grant number 451-03-34/2026-03/200107, dated February 05. 2026].

References

[1] DHILLON,B.S.: *Design reliability, fundamentals and applications*, Department of Mechanical Engineering, University of Ottawa, Ottawa, Ontario, Canada, CRC Press LLC, 1999.

[2] BERTSCHE,B.: *Reliability in automotive and mechanical engineering.* VDI-Buch, Springer-Verlag, Berlin Heidelberg, 2008.

[3] ERICSON,C: *Fault Tree Analysis – A History*, in: *Proceeding of the 17th International System Safety Conference*, 1999, http://www.fault-tree.net/ papers /ericson-fta-history.pdf. Accessed: 17. 01. 2010.

[4] CATIC,D., M. GASIC,M., SAVKOVIC,M., GLISOVIC,J.: *Fault tree analysis of hydraulic power-steering system*, International Journal of Vehicle Design, Inderscience Publishers, United Kingdom, Vol. 64, No 1, pp. 26-45, 2014.

[5] STAMATIS,D.H.: *Failure mode and effect analysis*, 2nd ed., American Society for Quality, Milwaukee, Wisconsin, 2003.

[6] PEETERS,J.F.W., BASTEN,R.J.I., TINGA,T.: *Improving failure analysis efficiency by combining FTA and FMEA in a recursive manner,* Reliability Engineering & System Safety, DOI: 10.1016/j.ress. 2017.11.024

[7] SHAFIEE,M., ENJEMA,E., KOLIOS,A.: *An integrated FTA-FMEA model for risk analysis of engineering systems*, Journal: Applied Sciences, DOI: 10.3390/app9061192, 2019.

[8] Military standard MIL-STD-1629A (1980). *Procedures for Performing a Failure Mode, Effects and Criticality Analysis* Department of Defense, Washington.

[9] ĆATIĆ,D.: *Development of accelerated experimental testing procedures for reliability assessment of systems in mechanical engineering*, Master'ş thesis, Faculty of mechanical engineering in Kragujevac, 1995.

[10] ĆATIĆ,D.: *Fault tree analysis of the causes of steering system failures in light commercial vehicles*, Statistical methods and quality system certification, OMO, Belgrade., (1994)., pp. 131-142.

[11] ĆATIĆ,D., JOVIČIĆ,S.: *Analysis of failures of the light industrial vehicle steering system by using the fault tree technique*, MVM - Volume 22, Number 1/2 - Kragujevac, (1996), pp. 14-29.

[12] ĆATIĆ,D., PAPIĆ,LJ.: *Ranking of steering system components in light commercial vehicles according to their criticality level*, Statistical methods and quality system certification, OMO, Belgrade., (1994), pp. 115-130.

[13] ĆATIĆ,D.: *Development and application of reliability theory methods*, Monograph, Faculty of mechanical engineering, Kragujevac, 2005.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

A QUALITY ASSURANCE FRAMEWORK FOR METAL ADDITIVE MANUFACTURING

SRDJAN ŽIVKOVIĆ
Military Technical Institute, Belgrade, srdjan.zivkovic@mod.gov.rs

SLOBODAN MALBAŠIĆ
Ministry of Defence – Defence Technologies Department, Belgrade, slobodan.malbasic@mod.gov.rs

MILOŠ STEPANOVIĆ
Military Technical Institute, Belgrade, hunter.milos@gmail.com

Abstract: *From a defense perspective, Additive Manufacturing (AM) has evolved from a rapid prototyping technology into a strategic capability and has become one of the key enablers of military modernization. Its applications include replenishing spare-part inventories for legacy platforms, enhancing supply chain agility, supporting forward operational logistics through on-site production of spare parts and rapid repairs, and improving overall operational readiness. Original Equipment Manufacturers (OEMs) and defense industries must rapidly adapt to AM technologies and comply with a comprehensive set of qualification and certification requirements to become qualified AM producers capable of manufacturing and delivering critical spare parts.*

The qualification of AM processes and parts is a highly challenging and demanding task, primarily defined by ***ISO/ASTM 52920:2023****, the international standard for quality assurance in additive manufacturing. Within the Quality Assurance (QA) framework, certification and qualification serve as essential tools for verifying and validating that manufacturing processes, parts and personnel comply with established standards and specified requirements.*

This paper provides a brief overview of the certification and qualification process for AM processes and components. Furthermore, it discusses the key challenges associated with conventional and AM-specific qualification approaches, as well as a novel qualification methodology currently being developed by the U.S. National Institute of Standards and Technology (NIST). Finally, the paper presents a basic framework for the qualification of AM parts intended for use in the armament and military equipment of the Serbian Armed Forces (SAF).

Keywords: *Quality Assurance, Additive Manufacturing, Qualification, Certification.*

1. INTRODUCTION

Quality standards, in the broadest sense, are established to ensure the consistent quality of industrial products over an extended period. When machine parts are manufactured using additive manufacturing (AM) technologies, it is essential to verify that they meet all specified technical and quality requirements. This requires that every stage of the production chain (Figure 1) be designed and controlled to ensure the consistent and repeatable quality of manufactured parts.

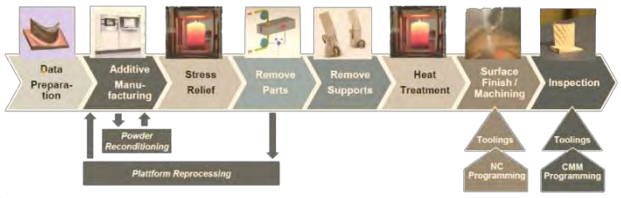


Figure 1. AM operations flow diagram

Quality Assurance (QA) is a comprehensive and systematic approach aimed at preventing defects and ensuring that a product or service fully complies with specified standards and user requirements before it reaches the end user.

Unlike Quality Control (QC), which focuses on detecting defects in finished products, QA encompasses the entire manufacturing process. It ensures process quality and stability during machine operation (in-process QA), as well as throughout the pre-process phase (input material verification and process preparation) and the post-process phase (inspection and testing of the finished part).

Quality Assurance (QA) uses Certification and Qualification as key tools to ensure and validate that processes meet set standards. The certification process is to assure the compliance with qualified parameters and standards. The qualification process is to demonstrate the ability to fulfil specified metal AM requirements, [1].

A specific aspect of quality assurance in additive manufacturing concerns production facilities and infrastructure, which must comply with stringent environmental control requirements. In addition, personnel involved in AM processes must possess appropriate qualifications and competencies.

The rapid adoption of metal additive manufacturing has introduced numerous technical challenges related to

Corresponding author: S. Živković;
E-mail: srdjan.zivkovic@mod.gov.rs

DOI: 10.5937/OTEH26145Z

ensuring product quality and confidence in component performance. Despite significant technological advances, concerns remain regarding whether metal AM parts can reliably replace components manufactured using conventional production methods, particularly in safety-critical applications. Nevertheless, additive manufacturing technologies are steadily becoming an integral part of modern manufacturing alongside conventional production technologies.

To achieve reliable and continuous quality management in industrial AM production, manufacturing equipment must be qualified in accordance with ISO/ASTM 52920:2023 [2], while programmers and machine operators must be qualified in accordance with the applicable personnel qualification standards [3,4]. Both qualifications are mandatory prerequisites for the certification of metal AM parts. Only after successful certification manufactured components (particularly those intended for load-bearing or mission-critical applications) can be integrated into military equipment.

Additive manufacturing is transforming defense procurement, from rapid prototyping and field repair to the production and deployment of qualified end-use components. Armed forces increasingly employ AM to enhance operational readiness by reducing supply chain vulnerabilities and enabling agile, on-demand, and localized manufacturing.

The primary AM technologies currently used in defense applications include, [5]:

- Polymer AM**,** particularly Material Extrusion (MEX) systems, which are widely adopted because of their relatively low acquisition cost, compact size, ease of operation, and plug-and-play capability.
- Metal Additive Manufacturing (MAM) technologies, including Laser Powder Bed Fusion (L-PBF), Directed Energy Deposition (DED), and Cold Spray. Although these technologies require significantly more complex infrastructure and logistics, they are rapidly advancing and are becoming increasingly important for the production of high-performance metal components.

The remainder of this paper is organized as follows. Section 2 provides an overview of the primary standards governing additive manufacturing. Section 3 reviews the conventional qualification process for parts manufactured using traditional production technologies and proposes a qualification new framework for AM components intended for military applications. Section 5 discusses personnel qualification requirements for Metal Additive Manufacturing. Finally, Section 6 outlines the infrastructure requirements for AM facilities and addresses the health and safety (H&S) aspects associated with metal additive manufacturing.

2. AM RELATED STANDARDS

ISO/TC 261 (International Organization for Standardization – Technical Committee) classifies AM standards into three levels, [6]: 1) *General standards:* which define general concepts, common requirements, or provisions applicable to most AM materials, processes, and applications, 2) *Category standards:* which specify requirements applicable to a particular material category (e.g., metal powders) or process category (e.g., Powder Bed Fusion), 3) *Specialized standards:* which define requirements for specific materials (e.g., aluminum alloy powders), specific processes (e.g., Material Extrusion using ABS), or application-specific sectors (e.g., aerospace, medical, or automotive industries).

To date, ISO/TC 261 has published 27 international standards**,** while 32 additional standards are currently under development, [7]**.** The large number of standards reflects the complexity and rapidly evolving nature of additive manufacturing technologies. As AM continues to mature and expand into new industrial sectors, further standardization activities are expected.

The development of AM standards is carried out jointly by ISO and ASTM International (formerly the American Society for Testing and Materials) through close cooperation between their respective technical committees. Such collaboration is essential to ensure that multidisciplinary expertise is incorporated into internationally recognized standards and to promote harmonized industrial practices worldwide.

The ISO/ASTM 52920:2023 standard establishes a comprehensive quality assurance framework that integrates system qualification, process monitoring, and part verification throughout the additive manufacturing process, Figure 2.

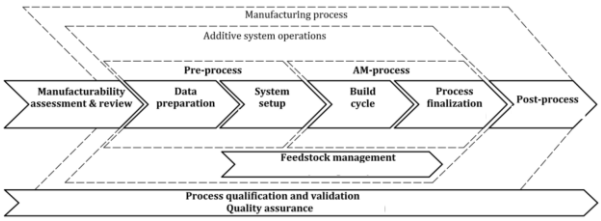


Figure 2. Overview of AM relevant processes in an AM production site

Furthermore, the standard defines the AM-related processes performed within an AM production facility and specifies the corresponding quality requirements for each process. The framework is organized into the following core technical areas, [2]: 1) Infrastructure of the part manufacturer; Manufacturability assessment and review; 2) Qualification of additive manufacturing system operations**,** focusing on machine qualification and risk assessment; 3) Quality-relevant process steps**,** addressing the control of production operations and material management; 4) Quality assurance**,** including personnel competence, final part verification, and the management of nonconforming products.

The minimum requirements for a Quality Management

System (QMS) are specified in ISO 9001. However, several industrial sectors have established sector-specific quality management standards that include additional requirements reflecting their particular regulatory and operational needs.

The most widely recognized examples include IATF 16949 for the automotive industry, ISO 13485 for medical devices, and EN/AS 9100 for the aerospace and defense industries. For organizations intending to supply AM components to the aerospace industry, certification according to EN/AS 9100 is generally considered a mandatory prerequisite, [8].

Based on the principles of ISO 9001, the EN/AS 9100 standard introduces additional requirements that are particularly relevant to aerospace and defense applications. For components manufactured using additive manufacturing technologies, these include: 1) Configuration management and process parameter control, 2) Complete material traceability throughout the manufacturing process, 3) Comprehensive risk management, 4) First Article Inspection (FAI), 5) Supplier qualification and control.

3. CERTIFICATION PROCESS DESCRIPTION AND PROPOSED FRAMEWORK

The certification of Metal Additive Manufacturing (MAM) parts is a complex and multidisciplinary process that typically involves several organizations and institutions. Outsourcing specific qualification activities to pre-qualified organizations can significantly reduce both the time and cost required for part certification while ensuring compliance with applicable standards.

Laser Powder Bed Fusion (L-PBF) is currently the most extensively researched additive manufacturing technology for aerospace applications and is therefore at the forefront of qualification and certification activities.

The National Aerospace and Defense Contractors Accreditation Program (Nadcap) was among the first certification bodies to establish a dedicated audit and accreditation program for Powder Bed Fusion technologies for the application in aviation, defense, and space programs. The Nadcap AM audit evaluates powder management, machine calibration, process control, and part qualification.

3.1 Certification procedure within the Serbian Ministry of Defense of the Republic of Serbia

Within the Serbian Ministry of Defense (MoD), the certification procedure for parts manufactured by the Original Equipment Manufacturer (OEM) is governed by the Regulation on Conformity Assessment in the Production of Armament and Military Equipment (AME) together with the corresponding Implementation Instructions, which provide detailed guidance for the execution of the conformity assessment procedure.

The conformity assessment is performed according to an approved Test Plan and Test Program by the relevant national institutions, including the Overhauling Facility Čačak, the Aeronautical Institute "Moma Stanojlović", the Technical Test Center (TTC), the Military Technical Institute (MTI), and, where appropriate, the relevant OEMs from the Serbian defence industry.

Upon successful completion of the certification activities, the OEM submits the required documentation to the Department for Defense Technology, including the test report, conformity assessment report, and evidence of compliance with applicable management system requirements, primarily the Quality Management System (QMS). Following a technical review of the submitted documentation, the Department issues the final Certificate of Conformity together with the approval for serial production.

3.2 Proposed framework for the certification of AM parts

The proposed framework is intended to optimize the qualification effort throughout the certification process while incorporating internationally recognized best practices for the industrial implementation of Additive Manufacturing.

At present, Serbia does not have a dedicated certification body accredited for the qualification of metal AM components. The institutions from MoD with the highest level of competence in this field are the Technical Test Center (TTC) and the Military Technical Institute (MTI). The proposed certification framework therefore should assume close cooperation between these two institutions. Although neither of these institutions individually possesses the competencies required to perform the complete qualification process, together they can provide comprehensive coverage of the required activities.

For additively manufactured components, conventional qualification procedures originally developed for traditional manufacturing technologies cannot be directly applied without modification. Consequently, a new qualification framework should clearly distinguish part verification from process and production-site qualification, including facility qualification.

The proposed qualification methodology consists of four sequential stages, Figure 3: 1) Process qualification, 2) Part qualification, 3) Functional testing under operational conditions, 4) Issuance of the Certificate of Conformity and production approval.

Process qualification verifies that materials, equipment, manufacturing parameters, and production workflows consistently produce components that satisfy the specified technical and quality requirements. The qualification methodology follows the internationally accepted three-stage approach consisting of Installation Qualification (IQ), Operational Qualification (OQ), and Performance Qualification (PQ), [9].

IQ confirms that the AM machine, software, auxiliary equipment, and supporting infrastructure are installed, calibrated, and maintained in accordance with the OEM's

specifications. This stage also verifies that environmental conditions, including temperature, humidity, ventilation, and safety systems, satisfy the required operating conditions.

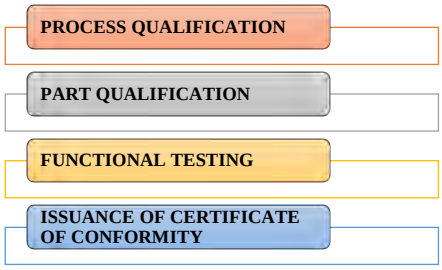


Figure 3. Framework proposal for am part qualification

OQ evaluates the critical process parameters, including laser power, scanning speed, layer thickness, hatch spacing, and other process variables. Qualification specimens are manufactured and tested to verify material properties, process capability, and manufacturing repeatability.

PQ demonstrates that the complete manufacturing process—from design and feedstock preparation to printing, post-processing, and inspection—can repeatedly produce conforming parts over multiple production runs. Qualification specimens are subjected to dimensional inspection, non-destructive testing (NDT), destructive testing, and mechanical property evaluation.

Regardless of the selected production model, whether components are manufactured in-house or supplied by an external supplier, the manufacturing organization must comply with the requirements of ISO/ASTM 52920:2023. These requirements are independent of the material system, manufacturing route, or final application. Major OEMs increasingly use this standard as the basis for auditing and approving contract manufacturers.

The certification of the manufacturing process and the qualification of an individual AM component are closely interconnected and cannot be treated as independent activities. An additively manufactured component cannot be qualified unless the complete manufacturing process has first been successfully qualified. Consequently, part qualification represents the final stage of the certified manufacturing process.

According to ISO/ASTM 52920:2023, part qualification includes the following key activities: 1) verification of the manufactured geometry against the approved CAD model, 2) strict configuration and change management to preserve the qualified ("frozen") manufacturing process, 3) validation through co-built witness coupons and First Article Inspection (FAI), including dimensional inspection, non-destructive testing, destructive testing, and mechanical property evaluation. The standard further requires requalification whenever any significant change is introduced into the manufacturing process, equipment, materials, or process parameters.

Upon completion of the qualification activities, the component is subjected to functional testing under representative operating conditions in accordance with the approved Test Plan and Test Program. Only after successful completion of all qualification and testing activities the Certificate of Conformity and the approval for production can be issued.

3.3 Future trends in C&Q

One of the major challenges facing Metal Additive Manufacturing (MAM) is achieving consistent process repeatability. Components manufactured on different machines (even when using the same process parameters and feedstock) may reveal variations in microstructure, mechanical properties, and dimensional accuracy. This lack of reproducibility significantly complicates and prolongs the qualification and certification process.

Another major challenge is the continued reliance on conventional qualification methodologies based primarily on extensive experimental testing. Although these approaches have proven effective for conventionally manufactured components, they are often unsuitable for additive manufacturing because they require large numbers of physical test specimens, resulting in high qualification costs and lengthy development cycles, particularly in highly regulated sectors such as the aerospace and defense industries.

To address these challenges, new qualification methodologies are currently being developed. One of the most significant initiatives is the Computational Materials for Qualification and Certification (CM4QC) program led by the National Institute of Standards and Technology (NIST), U.S. Department of Commerce. The objective of this program is to integrate validated computational models into the qualification process, enabling the prediction of component properties and performance while substantially reducing the dependence on costly physical testing, [10].

This model-based qualification approach relies on high-fidelity numerical simulations capable of predicting the influence of manufacturing parameters on the final characteristics of additively manufactured components. For such simulations to be accepted by certification authorities, their predictive capability must first be experimentally validated. To support this objective, NIST has established the Additive Manufacturing Benchmark Test Series (AM-Bench), an internationally recognized benchmark platform for validating computational models. AM-Bench consists of carefully designed experiments performed under strictly controlled conditions, generating high-quality reference datasets that enable researchers to evaluate and compare the accuracy of simulation tools worldwide.

Consequently, the AM-Bench datasets constitute the scientific foundation for establishing confidence in computational models and are expected to play a key role in the future implementation of model-based qualification and certification of additively manufactured components for aerospace, defense, and other safety-critical industries.

4. MAM PERSONNEL QUALIFICATION

Qualified personnel represent one of the most critical resources in additive manufacturing (AM), as in all advanced manufacturing technologies. Similar to other industrial fields, such as welding, the qualification of AM designers and machine operators requires a structured combination of theoretical education and practical training. Practical training is conducted under the supervision of qualified personnel and is followed by a formal qualification assessment consisting of written, oral, and practical examinations. Successful candidates receive certification authorizing them to perform specific AM-related tasks.

In addition to machine operation, other activities associated with additive manufacturing (including post-processing, machining, destructive testing, and non-destructive testing (NDT)) also require appropriately qualified personnel in accordance with the applicable qualification standards, [3].

4.1 Qualification of Operators for PBF-LB

The qualification of operators for Laser Powder Bed Fusion (PBF-LB) systems shall be performed in accordance with a documented qualification procedure or manufacturing plan. Prior to qualification, candidates shall complete the prescribed theoretical and practical training specified by the relevant qualification standard, [4].

The practical examination demonstrates that the operator possesses the knowledge and practical skills required to independently perform all manufacturing activities associated with a specific production task.

Upon successful completion of the qualification process, the operator shall be competent to understand and perform the following activities: 1) principles and operational aspects of the selected AM technology, 2) feedstock handling, storage, and material management, 3) AM system setup, calibration, and preparation, 4) build preparation, process monitoring, and manufacturing operations, 5) post-processing activities, including part removal, cleaning, and inspection.

4.2 Tasks and Competencies of the AM Designer

An **AM Designer** is a professional who applies knowledge of additive manufacturing process capabilities to develop optimized geometrical, functional, and manufacturing-oriented component designs for metal additive manufacturing, without necessarily being responsible for the execution of the manufacturing process [3].

An AM designer should possess both general manufacturing knowledge and specialized expertise in additive manufacturing technologies. In addition, the designer should be proficient in Computer-Aided Design (CAD), Computer-Aided Manufacturing (CAM), and the software tools required to support the complete AM workflow.

The wide variety of engineering challenges associated with metal additive manufacturing directly influences the competencies required of an AM designer. These include occupational health and safety (OH&S) considerations, material behavior, process limitations, geometric inspection, qualification testing, post-processing, and manufacturing economics.

Since Metal Additive Manufacturing (MAM) technologies differ significantly in terms of process characteristics and manufacturing constraints, the designer must understand how these differences influence component design. Typical design considerations include machining allowances for post-processing, build orientation, support structure design, residual stress mitigation, and the selection, location, and orientation of qualification specimens.

An AM designer should be able to understand and perform the following tasks: 1) understand the fundamentals of additive manufacturing technologies and their operating principles, 2) apply Design for Additive Manufacturing (DfAM) principles, 3) develop component designs optimized for additive manufacturing, 4) perform pre-processing and build data preparation, 5) define process control requirements during manufacturing, 6) identify appropriate post-processing methods based on functional requirements, 7) evaluate manufacturing costs while considering technical and quality requirements, according to [3].

5. INFRASTRUCTURE AND OH&S CHALLENGES

5.1 Infrastructure challenges

The highest level of qualification for an organization involved in Additive Manufacturing (AM) extends beyond the qualification of manufacturing processes and includes the organization's overall quality infrastructure. The scope of qualification depends on the company's activities and production portfolio.

For example, the MTI has implemented an Integrated Quality Management System (IQMS) comprising certification according to several internationally recognized standards, including ISO 9001 (Quality Management System), ISO 14001 (Environmental Management System), ISO 45001 (Occupational Health and Safety Management System - OH&S), ISO/IEC 27001 (Information Security Management System), and accreditation of testing laboratories in accordance with ISO/IEC 17025.

In addition to implementing an Integrated Quality Management System, the introduction of advanced Metal Additive Manufacturing technologies, such as Laser Powder Bed Fusion (PBF-LB/M), requires additional qualification procedures. In particular, qualification of the production facility in which the AM equipment is installed is mandatory.

A qualified AM production facility should include dedicated areas for design and engineering activities, production, metrology, post-processing, material handling and storage, quality inspection, and supporting activities

required throughout the AM workflow illustrated in Figure 1.

Ultimately, a qualified AM production facility consists of qualified personnel, qualified equipment, and qualified manufacturing processes operating within an integrated quality management framework to consistently produce components that satisfy all specified requirements, [9].

5.2 OH&S challenges

The use of metallic powders in additive manufacturing introduces a number of new occupational health and safety challenges. Although many metallic alloys employed in AM present relatively low health risks, fine aluminum and titanium powders represent significant fire and explosion hazards because of their high reactivity and large specific surface area. As metal additive manufacturing continues to expand across industrial sectors, organizations must identify, assess, and mitigate these hazards through systematic risk management.

Risk mitigation should follow the internationally accepted **Hierarchy of Controls**, which consists of five levels: 1) Elimination, 2) Substitution, 3) Engineering Controls, 4) Administrative Controls, 5) Personal Protective Equipment (PPE).

Before implementing Metal Additive Manufacturing technologies, organizations should perform a comprehensive occupational health and safety assessment covering all stages of powder handling, machine operation, post-processing, maintenance, waste management, and emergency response procedures, [11].

6. CONCLUSION

The qualification and certification of metal components manufactured by AM represent a new and strategically important field for the Republic of Serbia, particularly in the context of defence applications. The implementation of internationally recognized qualification principles is essential for ensuring the reliability, repeatability, and traceability of AM processes and for enabling the safe integration of additively manufactured components into military equipment.

This paper has presented an overview of the qualification and certification requirements for Metal Additive Manufacturing (MAM), with particular emphasis on the requirements defined in ISO/ASTM 52920:2023**.** The proposed framework aligns the requirements of international AM standards with the existing national conformity assessment system and identifies the complementary roles of the Technical Test Center (TTC) and the Military Technical Institute (MTI) in establishing a future national certification capability.

The paper also discussed the differences between conventional and additive manufacturing qualification approaches and highlighted emerging trends, including model-based qualification initiatives currently being developed by NIST.

Future research will focus on validating the proposed qualification framework through representative case studies, establishing standardized testing procedures for AM components intended for military applications, and supporting the development of an accredited national certification capability.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No 451-03-34/2026-03/200325).

References

[1] CHEN, Z., HAN, C., GAO, M., KANDUKURI, S. Y., & ZHOU, K.: *A review on qualification and certification for metal additive manufacturing. Virtual and Physical Prototyping*, 17(2), 382–405, 2022. https://doi.org/10.1080/17452759.2021.2018938.

[2] ISO/ASTM 52937:2024: Additive manufacturing of metals — Qualification principles — Tasks and related skills for AM, https://www.iso.org/standards.html.

[3] ISO/ASTM 52920:2023: Additive manufacturing — Qualification principles — Requirements for industrial additive manufacturing processes and production sites, https://www.iso.org/standards.html.

[4] ISO/ASTM 52926-2:2023: *Additive manufacturing of metals — Qualification principles — Part 2: Qualification of operators for PBF-LB/M* (Powder Bed Fusion – Laser Beam with Metals), https://www.iso.org/standards.html.

[5] AMPOWER: *Additive Manufacturing in Defense NOVEMBER 2025,* https://ampower.eu/insights/additive-manufacturing-defense/

[6] SEIFI, M., GORELIK, M., WALLER, J., HRABE, N., SHAMSAEI, N., DANIEWICZ, S., & LEWANDOWSKI, J.J.: *Progress towards metal additive manufacturing standardization to support qualification and certification. JOM*, *69*(3),439–455, 2017. https://doi.org/10.1007/s11837-017-2265-2

[7] LJUBOJEVIĆ P., LAZOVIĆ T., ĆIRIĆ-KOSTIĆ S.; *Additive manufacturing - a view through the prism of standardization*, Proceedings of XI International Conference "Heavy Machinery-HM 2023", Vrnjačka Banja, 21 – 24 June 2023. https://machinery.mas.bg.ac.rs/bitstream/id/17399/bitstream_17399.pdf

[8] AMPOWER: *Quality in Additive Manufacturing*, *Best practices for quality management in Additive Manufacturing*, VOLUME 4 NOVEMBER 2018, https://ampower.eu/insights/quality-additive-manufacturing/

[9] ISO/ASTM TS 52930:2021: Additive manufacturing — Qualification principles — Installation, operation and performance (IQ/OQ/PQ) of PBF-LB equipment (Technical Specification), https://www.iso.org/standards.html.

[10] NIST (National Institute of Standards and

Technology, U.S. Department of Commerce**):** *Computational materials for qualification and certification.* https://www.nist.gov/programs-projects/cm4qc.

[11] ŽIVKOVIĆ S., MALBAŠIĆ S., STEPANOVIĆ M., KASALICA V., STOJADINOVIĆ S.: *Metal Additive Manufacturing OHS Challenges and Guidelines in Implementation*; Proceedings XLV JUPITER Conference, pp. 2.121-2.128; ISBN 978-86-6060-250-5, Belgrade 9-10. Jun 2026.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ELECTRO-MECHANICAL FUZES FOR UAV LOITERING MUNITION WARHEADS: A LAYERED SAFETY AND STANDARDIZATION FRAMEWORK

MARINKO ALEKSIĆ
HI POLIEX AD, Research and Development Center, Berane, Montenegro Email: sigurnostnamoru@yahoo.com

NIKOLA NIKOLIĆ
HI POLIEX AD, Research and Development Center, Berane, Montenegro Email: nikolanikolic953@gmail.com

MARIJA BOROVIĆ
HI POLIEX AD, Research and Development Center, Berane, Montenegro Email: marija@poliex.me

DEJAN MARSENIĆ
HI POLIEX AD, Research and Development Center, Berane, Montenegro Email: dejan.marso@gmail.com

DANIJELA KOVAČ
Kovački Centar d.o.o., Valjevo, OJ Mladenovac, Mladenovac, Serbia Email: danijela.kovac@kovacki-centar.com

NEMANJA VUJOVIĆ
HI POLIEX AD, Research and Development Center, Berane, Montenegro Email: vujovicnemanja2222@gmail.com

Abstract: *The character of modern warfare has been significantly affected by the widespread use of unmanned aerial vehicles and loitering munitions. Their use as weapon carriers introduces safety challenges that differ from those encountered in conventional artillery, mortar and rocket ammunition. Traditional fuzing systems commonly rely on setback acceleration, rotational motion, centrifugal effects or a combination of such mechanical conditions to achieve safe arming after launch or firing. In contrast, many UAV-based configurations do not provide comparable mechanical arming environments. This creates a need for a dedicated safety approach for warheads integrated with unmanned platforms, particularly when electrical or electromechanical fuzes are used. This paper reviews the safety rationale behind existing fuze-related standards, identifies limitations of direct transfer of conventional fuze safety assumptions to UAV-borne warheads, and proposes a layered safety concept. The proposed framework combines administrative control, positive physical safety, platform-state verification, electrical and command interlocks, built-in test functions, fault response and abort-related safe-state logic. An illustrative Electronic Safe-and-Arm Device (ESAD) state-transition model is presented as a verification-oriented example. The paper does not propose a detailed fuze design, but provides a conceptual, system-safety and standardization-oriented framework for future assessment of electromechanical fuzing systems used with loitering munition warheads.*

Keywords: *loitering munition, electromechanical fuze, ESAD, safety analysis, UAV-borne warhead, standardization, built-in test, system safety.*

1. INTRODUCTION

Fuzes and safety-and-arming devices are among the most safety-critical subsystems of any munition. Their function is not limited to initiating a warhead at the required moment. They must also prevent unintended initiation during manufacture, transport, storage, handling, integration, launch preparation and all pre-arm conditions.

In artillery and mortar ammunition, arming is commonly linked to launch conditions. Setback acceleration generated during firing may release or enable internal mechanical elements only after the munition has left the weapon system. In spin-stabilized ammunition, rotational motion and centrifugal effects may also contribute to the arming sequence. Similar principles are applied in many rocket and missile systems, where launch acceleration or other launch-related parameters provide a clear physical distinction between a safe handling state and an armed flight state.

The increasing use of unmanned aerial vehicles and loitering munitions changes this safety context. A UAV is an aircraft-like platform that may be assembled, loaded, powered, checked, launched, flown, aborted or recovered under operational conditions that differ substantially from those of classical ammunition. In many cases, the acceleration profile is too low to be used as a reliable arming parameter. Rotational arming is also generally not applicable.

This paper discusses why existing fuze safety concepts

Corresponding author:857 N. Nikolić;
E-mail: nikolanikolic953@gmail.com
DOI: 10.5937/OTEH26146A

cannot be transferred directly to UAV-borne warhead applications and proposes a layered framework for safe integration, handling, pre-launch control, mission-state management and verification of warheads with electromechanical fuzes.

The central argument of the paper is that the safety of an electromechanical fuze for a UAV-borne warhead should not be assessed as an isolated component problem. It should be assessed as a platform-integrated system safety problem, where the fuze, UAV platform, electrical interfaces, command logic, operator procedures and mission state form one safety-relevant system.

2. METHODOLOGICAL APPROACH

This paper applies a conceptual system-safety methodology based on four complementary steps.

First, the paper compares the conventional fuze arming environment with the mission and integration environment of UAV-borne warheads. This comparison is used to identify safety assumptions that may no longer be valid when a conventional warhead or fuze concept is adapted to an unmanned platform.

Second, the paper interprets existing fuze and munition safety standards from the perspective of UAV integration. The objective is not to replace established standards, but to determine which safety principles remain directly applicable and which principles require additional interpretation in the UAV-borne warhead context.

Third, the paper decomposes the lifecycle of a UAV-borne warhead into storage, transport, integration, pre-launch preparation, flight, abort and post-mission handling. Each lifecycle block is considered in relation to credible unsafe states and the safety layers required to prevent or mitigate them.

Fourth, the paper uses a simplified ESAD state-transition model to illustrate how safety-related states and mission-readiness states may be coordinated. The model is verification-oriented. It is not intended to represent a complete fuze design, electrical circuit, firing chain or detailed implementation.

This methodological approach is suitable for a conference-level standardization and safety analysis paper because it combines comparative analysis, lifecycle reasoning, hazard-barrier mapping and state-transition modelling without disclosing design-sensitive implementation details.

3. CONVENTIONAL FUZE SAFETY PRINCIPLES AND LIMITATIONS

Fuze safety standards are built around the principle that the i train must remain interrupted or otherwise controlled until a valid arming environment has been achieved. This principle is supported by mechanical barriers, environmental discrimination, arming delay, safety distances, qualification tests and verification of safe handling conditions. Standards such as MIL-STD-1316, MIL-STD-331 and related NATO ordnance publications provide a strong foundation for this approach.

The limitation in UAV applications arises from the fact that the expected arming environment may be absent, weak or fundamentally different from that of the original munition. In gun-launched ammunition, firing creates a short and intense acceleration environment. In spin-stabilized ammunition, rotation provides an additional physical input. By contrast, UAVs often experience gradual acceleration, low launch shock and relatively smooth flight profiles.

A second limitation is related to mission variability. A projectile normally follows an irreversible trajectory after firing. A UAV-based system may remain in flight for an extended period, loiter before engagement, receive operator commands, abort a mission or return to a predefined area.

A third limitation is the integration of the fuze into an electrically powered platform, with dependencies on power supply, avionics, data links, command logic and system interfaces. In such a system, electrical safety cannot be treated only as the presence or absence of power. It must also include power validity, command validity, logic enable conditions, fault detection, reset behavior, loss-of-power response, electromagnetic immunity and defined safe-state behavior.

Therefore, the direct adaptation of a conventional fuze or safety-and-arming philosophy to a UAV-borne warhead can be unsafe if the original assumptions about launch environment, trajectory, operator interaction and post-launch irreversibility are no longer valid.

Table 1. Conceptual comparison of conventional and UAV-borne warhead safety environments

Aspect	Conventional munition	UAV-borne warhead
Launch event	Strong firing or launch signature	Low or variable launch signature
Rotation	May be present in spin-stabilized projectiles	Usually absent
Mission profile	Mostly irreversible after firing	May include loiter, delayed engagement, abort or recovery
Power and interfaces	Mainly fuze-contained or weapon-system related	Platform power and mission-state dependencies
Operator interaction	Mainly before firing	May continue before and during mission
Environmental exposure	Defined by munition class	Defined by air vehicle and carrier context

Safety assumption	Valid arming environment is physically distinct from handling state	Valid arming environment must be defined by combined platform, command and mission-state conditions

4. PROPOSED LAYERED SAFETY CONCEPT

A suitable safety concept for electromechanical fuzes in UAV-borne warhead applications should combine procedural, mechanical, electrical and system-level controls. The purpose is not to replace existing fuze safety standards, but to adapt their safety philosophy to the operational environment of UAV platforms.

The proposed concept is based on the principle that no single safety layer should be treated as sufficient on its own. Safety should be achieved through a controlled sequence of independent and mutually reinforcing barriers, covering human procedures, physical safety features, platform status, electrical command logic, fault detection and response to abnormal mission conditions.

Figure 1. Proposed layered safety concept for UAV-borne electromechanical fuzes.

The first layer is administrative and procedural control. It defines responsibility, authorization, configuration control and the conditions under which the warhead may be handled, integrated or prepared for mission use.

The second layer is positive physical safety. This layer provides an independent non-electrical barrier and is particularly important because electrical systems may be affected by power faults, software faults, electromagnetic interference, electrostatic discharge or incorrect interface states.

The third layer is platform-state verification. The warhead and fuze should not be allowed to progress toward an unsafe state unless the UAV platform is in a valid mission state and all defined preconditions are satisfied.

The fourth layer consists of electrical and command interlocks. These interlocks should support controlled application of power, valid logic enable states, authorized command paths and rejection of invalid or ambiguous command conditions.

The fifth layer includes built-in test and fault response. Built-in test should support detection of safety-critical faults before launch and, where applicable, during mission execution. Fault response should be biased toward a defined safe state.

The sixth layer includes abort and safe-state logic. This layer is especially important for UAV-borne warheads because the mission may be interrupted, delayed or aborted. Unlike conventional projectiles, a UAV-based system may require defined transitions back from an unsafe intermediate state to a controlled safe or less unsafe state.

Table 2. Proposed safety layers for UAV-borne electromechanical fuzes

Safety layer	Function	Safety role
Administrative and procedural control	Responsibility, authorization and configuration control	Controls human-factor and organizational risk
Positive physical safety	Independent safety barrier	Provides non-electrical protection against premature unsafe transition
Platform-state verification	Valid platform and mission state	Prevents arming outside approved platform states
Electrical and command interlocks	Controlled power, logic enable and valid command conditions	Mitigates electrical, interface or command faults
Built-in test and fault response	Detection of safety-critical faults	Supports fail-safe or safe-state behavior
Abort and safe-state logic	Response to abnormal mission outcomes	Controls residual risk during abort, delay or post-mission handling

5. SAFETY RISK ANALYSIS FRAMEWORK

The safety analysis should cover the complete lifecycle of a UAV-borne warhead, from storage and transport to integration, pre-launch preparation, flight, mission abort and post-mission handling. Each phase should be assessed against credible initiating events, possible unsafe states and the safety layers required to prevent or mitigate them.

The proposed analysis is structured around three questions: which unsafe state may occur, in which lifecycle block it may occur and which safety layer provides the primary barrier. This approach keeps the assessment system-oriented and avoids treating the fuze as an isolated component. The safety of an electromechanical fuze depends on its interaction with the UAV platform, electrical interfaces, command logic, operator procedures and environmental exposure.
For conference-level assessment, a compressed hazard-barrier-verification structure is useful because it links the unsafe state to the lifecycle phase, the primary safety barrier and the type of evidence needed for verification.

Table 3. Lifecycle structure of safety risk analysis

Lifecycle block	Main safety concern	Control focus
Storage and transport	Loss of safe configuration	Physical safety, packaging and configuration status
Integration and pre-launch	Incorrect assembly or premature transition	Procedure, visible status, platform checks and BIT
Flight and mission execution	Unintended command, fault or unsafe transition	Interlocks, authorization, BIT and state monitoring
Abort and post-mission handling	Unsafe residual state	Safe-state logic and status confirmation

Table 4. Example hazard-barrier-verification matrix for UAV-borne electromechanical fuzing safety

Hazard scenario	Lifecycle phase	Primary safety barrier	Verification evidence
Premature transition toward arm-ready state	Integration / pre-launch	Platform-state verification and procedural control	Pre-launch checklist, platform status confirmation and BIT result
Invalid or unintended electrical command	Pre-launch / flight	Electrical and command interlocks	Interface test, command validity check and logic verification
Loss of power or voltage instability	Flight / abort	Safe-state logic and power validity monitoring	Power interruption test and reset behavior verification
Ambiguous state after mission abort	Abort / post-mission handling	Abort logic and status indication	Safe-state confirmation and operator procedure
Electromagnetic or electrostatic disturbance affecting safety-critical logic	Integration / flight	EMC/ESD qualification and fault response	MIL-STD-461-oriented EMC assessment and ESD test evidence
Incorrect configuration during integration	Integration	Administrative control and physical safety	Configuration checklist and independent status confirmation
Built-in test detects safety-critical fault	Pre-launch / flight	BIT and fault response	Fault-insertion or diagnostic response verification
Environmental exposure affects electrical or mechanical integrity	Storage / transport / flight	Environmental qualification	Temperature, vibration, shock, humidity or salt-fog test evidence

6. ILLUSTRATIVE ESAD STATE-TRANSITION EXAMPLE

In order to complement the proposed layered safety concept with a concrete technical illustration, Fig. 2 presents a simplified ESAD state-transition example influenced by the loitering-munition mission state. In this paper, ESAD is used in the sense of an Electronic Safe-and-Arm Device. The figure is not intended to represent a complete system specification. It is a reduced verification-oriented model showing how safety-related states and mission-readiness states may be coordinated.

The model should be understood as a state-management example, not as a circuit diagram or detailed fuze design. It separates three categories of states: safe-for-handling states, controlled unsafe intermediate states and terminal unsafe states. This distinction is important because UAV-borne warheads may remain associated with the platform for a longer period and may be subject to mission changes, operator commands or abort conditions.

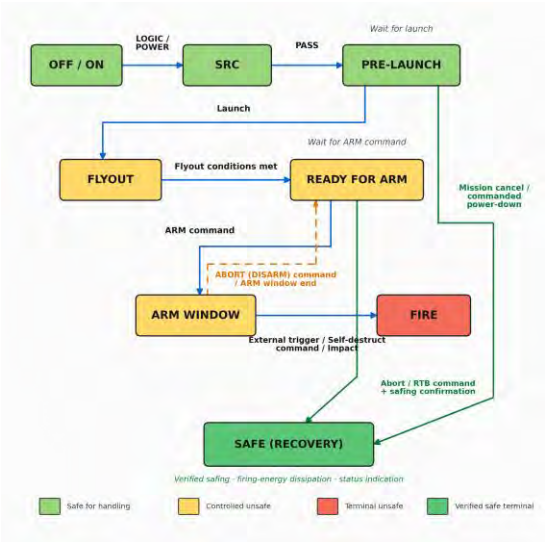


Figure 2. Illustrative ESAD state-transition example influenced by loitering-munition mission states.

The example begins with the OFF state, followed by POWER ON / LOGIC ENABLED. The system then enters FSC, representing the Final System Check of the complete platform-warhead configuration. The purpose of FSC is to confirm that required readiness and safety conditions are satisfied before the system can proceed to the PRE-LAUNCH state.

After successful completion of the final system check, indicated by PASS, the system enters the PRE-LAUNCH state. In this state, the system remains safe for handling and waits for the launch condition. Following the Launch transition, the system proceeds to FLYOUT. Once the required flyout conditions are met, the logic transitions to READY FOR ARM.

The READY FOR ARM state represents a controlled intermediate condition. In this condition, the warhead with electromechanical fuze is no longer treated as a normal handling state, but it is not yet within the final arming window. The transition from READY FOR ARM to ARM WINDOW requires an explicit ARM command and valid mission-state conditions. The ARM WINDOW state is therefore treated as a restricted state in which terminal transition to FIRE may occur only under defined event conditions represented in the simplified model as External Trigger, Self-Destruct Command or Impact.

The return path from ARM WINDOW to READY FOR ARM, initiated by an ABORT / DISARM command or by ARM Window End, illustrates the importance of controlled state management. Unlike conventional projectiles, a UAV-borne warhead may be associated with mission interruption, delayed engagement or abort-related conditions. For this reason, the state logic should support clear differentiation between safe handling states, controlled unsafe transition states and terminal unsafe states.

This example demonstrates how safety may be implemented not only through a fuze mechanism, but also through coordinated interaction between ESAD logic and the mission state. In this way, the state-transition model supports risk reduction by preventing premature arming, requiring readiness verification before launch, separating flyout from arming authorization and providing defined paths for abort-related state control.

For verification purposes, each state and each transition should have defined entry conditions, exit conditions, permitted commands, prohibited commands, fault response and status indication. In particular, the system safety case should address loss of power, voltage instability, logic reset, invalid command sequence, communication fault, electromagnetic disturbance and operator-visible status after abort.

7. STANDARDIZATION CONSIDERATIONS

Existing standards provide a strong foundation for fuze safety, environmental qualification and testing. MIL-STD-1316 defines safety criteria for fuze design, while MIL-STD-331 addresses environmental and performance testing of fuzes, ignition safety devices and related components. MIL-STD-810 provides a broader framework for environmental engineering considerations and laboratory tests. MIL-STD-461 defines requirements for the control of electromagnetic interference characteristics of subsystems and equipment. These standards are directly relevant to electromechanical fuzes, but their application to UAV-borne warheads requires careful interpretation.

The main standardization challenge is not the absence of general safety principles. The challenge is the lack of a dedicated framework for applying those principles to warheads integrated with unmanned platforms. These systems combine characteristics of ammunition, aircraft systems, electrical subsystems and mission-controlled platforms. Their safety assessment should therefore integrate ordnance safety, air vehicle integration, electrical interface control, environmental qualification and operational safety procedures.

A future standard or guidance document should define minimum expectations for safe-state definition, positive physical safety, platform-state criteria, electrical and command interlocks, built-in test requirements, critical fault response, operator status indication, electromagnetic compatibility and immunity, electrostatic discharge protection, loss-of-power and reset behavior, abort logic, post-mission safe-state confirmation, and verification and validation evidence.

Environmental requirements should be treated as safety-relevant. UAV-borne systems may be exposed to vibration, shock, temperature variation, humidity, salt fog, electrostatic discharge and electromagnetic effects. These conditions may influence electrical interfaces, mechanical integrity and the reliability of safety-critical functions. Environmental testing should therefore be part of the overall safety case, not only a durability issue.

The standardization framework should also recognize the difference between design qualification and operational safety. A fuze or ESAD may satisfy component-level requirements and still present system-level risks if the

platform interface, command logic, operator procedure or abort state is not adequately controlled. For this reason, standardization should require a system-level safety case covering both component qualification and platform integration.

8. DISCUSSION

The proposed layered concept has two main advantages. First, it preserves the safety philosophy of existing fuze standards by maintaining the principle that no hazardous state should be reached without valid and verified conditions. Second, it adapts this philosophy to the specific operational reality of UAV-borne warheads, where launch signature, rotation and irreversible ballistic trajectory cannot be assumed.

The most important implication is that electromechanical fuzing safety should be treated as a system property. It is not sufficient to qualify only the fuze hardware. The UAV platform, power supply, electrical interfaces, command architecture, mission logic, operator procedures and environmental exposure must also be included in the safety assessment.

A second implication is that built-in test should not be treated merely as a maintenance or readiness feature. In UAV-borne warhead applications, BIT may become part of the safety argument because it can prevent continuation of the mission sequence when a safety-critical fault is detected. However, BIT itself must be carefully verified, because a false pass, false status indication or ambiguous diagnostic result may create additional risk.

A third implication is that abort and post-mission handling require special attention. Conventional ammunition safety logic often assumes an irreversible post-launch trajectory. UAV-borne systems may not follow this assumption. Therefore, standards should explicitly address mission interruption, delayed engagement, return-to-area conditions and operator confirmation of safe state after abort.

The proposed framework is intentionally conservative. It does not rely on a single command, a single sensor, a single software condition or a single operator action. Instead, it requires mutually reinforcing safety layers and objective verification evidence.

9. CONCLUSION

The integration of warheads and electromechanical fuzes into UAVs and loitering munitions creates a distinct safety problem. Classical fuze arming principles rely heavily on launch acceleration, setback forces or rotational effects. These parameters are often absent or insufficient in UAV applications. As a result, the direct adaptation of conventional ammunition to unmanned platforms may create unacceptable safety risks if the original safety assumptions are no longer valid.

This paper proposes a layered safety concept for warheads with electromechanical fuzes used in UAV-borne applications. The proposed concept combines administrative control, positive physical safety, platform-state verification, electrical interlocking, built-in test functions, fault response and abort-related safe-state logic. The illustrative ESAD state-transition example shows how these principles can be translated into a verification-oriented state model.

The principal contribution of the paper is the proposal that UAV-borne warhead fuzing safety should be assessed as a platform-integrated system safety problem, rather than as a direct adaptation of conventional fuze qualification logic. Further work should include detailed hazard analysis, verification and validation methods, test evidence structure and development of operational safety protocols aligned with established munition safety principles.

Acknowledgements

This research was financially and technically supported by HI Poliex AD, Research and Development Center, Berane, Montenegro.

References

[1] MIL-STD-1316F, Fuze Design, Safety Criteria For, Department of Defense, 2017.

[2] MIL-STD-331D, Department of Defense Test Method Standard: Fuzes, Ignition Safety Devices and Other Related Components, Environmental and Performance Tests For, Department of Defense, 2017.

[3] MIL-STD-810G, Department of Defense Test Method Standard: Environmental Engineering Considerations and Laboratory Tests, Department of Defense, 2008.

[4] MIL-STD-461G, Requirements for the Control of Electromagnetic Interference Characteristics of Subsystems and Equipment, Department of Defense, 2015.

[5] NATO AOP-4187, Fuzing Systems, NATO Standardization Office.

[6] NATO AOP-15, Guidance on the Assessment of the Safety and Suitability for Service of Non-Nuclear Munitions for NATO Armed Forces, NATO Standardization Office.

[7] NATO STANAG 4187, Fuzing Systems, NATO Standardization Office.

[8] MIL-STD-882E, System Safety, Department of Defense, 2012.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CLIMATE CHANGES AND ITS IMPACTS ON WATER RESOURCES, CASE STUDY CRNICA RIVER BASIN

ALEKSANDAR TANASKOVIĆ
University of Belgrade, Faculty of Mining and Geology, Belgrade, Serbia, aleksandar.tanaskovic@rgf.bg.ac.rs, ORCID: 0009-0005-2668-3588

VESNA RISTIĆ VAKANJAC
University of Belgrade, Faculty of Mining and Geology, Belgrade, Serbia, and full member of IRASA, vesna.ristic@rgf.bg.ac.rs, ORCID: 0000-0002-2040-8252

VELJKO MARINOVIĆ
University of Belgrade, Faculty of Mining and Geology, Belgrade, Serbia, +381604404345, veljko.marinovic@rgf.bg.ac.rs, ORCID: 0000-0003-4749-7053

BORIS VAKANJAC
Metropolitan University, Faculty of Applied Ecology Futura, and full member of IRASA, Belgrade, Serbia, borivac@gmail.com, ORCID 0000-0002-9429-6943

HRISTINA PETROVA
PhD Student of University of Belgrade, Faculty of Mining and Geology; Belgrade, Serbia, E-mails: g808-23@dok.rgf.bg.ac.rs ; Geolesnovo DOO Skopje, Republic of North Macedonia, ORCID: 0009-0004-6697-0920

Abstract: *From the perspective of water supply, the inhabitants of the Republic of Serbia largely (over 75%) meet their needs from groundwater. The estimated reserves of groundwater across the entire territory of Serbia amount to about 68 m³/s, of which only one third is currently utilized, i.e., around 23 m³/s of the total available potential of this strategic resource. These reserves are not uniformly distributed throughout Serbia, and therefore they are not equally accessible to all inhabitants. Considering that climate change has a significant impact on this resource, the paper will present a comparative analysis of direct (surface), base (groundwater) runoff, as well as the baseflow index analyzed over different time periods. As a pilot area, the Crnica River basin was selected, which belongs to river courses with one of the longest periods of water level and discharge monitoring. Today, about 40,000 inhabitants of the Municipality of Paraćin and part of the population of the settlement of Ćuprija use groundwater from this basin for water supply purposes. Monitoring of quantitative parameters on this river began in 1925 and continues to this day. For the purpose of analyzing the impact of climate change on the water regime of this river, three different 30-year time periods were distinguished: Period I (1931–1960), Period II (1961–1990), and Period III (1991–2020).*

Keywords: *direct runoff, baseflow, baseflow index, climate change, Crnica River*

1. INTRODUCTION

The Crnica catchment is in the central part of Serbia. It is delineated by the valley of the Velika Morava to the west, the Sisevac Basin, Kučaj and Samanjic to the east, and the catchments of the Ravanica to the north and the Jovanica River to the south. In its source, upper part (eastern section), the catchment has a wide, arcuate shape, while in the lower part (western section), towards the Velika Morava into which the Crnica River flows, the catchment narrows wedge-like. In the observed area of the Crnica River catchment, from a hydrogeological standpoint, all types of aquifers are present, among which the karst type aquifer plays a more significant role (Figure 1).

The dominant karst aquifer type in the upper part of the Crnica River catchment (yellow area in Figure 1), owing to cavernous porosity encompassing systems of fractures, caverns, and cave channels, enables the formation of considerable reserves of groundwater. The ability of the terrain to retain water infiltrated into the catchment and subsequently return it to the river flow represents a key hydrogeological characteristic, which largely determines the runoff regime during dry periods. The process of groundwater depletion is described by the shape of the recession limb of the hydrograph, i.e., the recession curve. The very character of the recession curve is determined by analyzing the falling limbs of a larger number of flood wave hydrographs. Accordingly, the total runoff is formed as the cumulative value of direct and base runoff. Direct runoff arises during intense precipitation and is present (dominant) mainly during flood waves. In contrast, base runoff occurs because of the river flow being recharged by groundwater (Prohaska, 2004).

Corresponding author: A. Tanasković
E-mail: aleksandar tanaskovic@rgf.bg.ac.rs

DOI: 10.5937/OTEH26147T

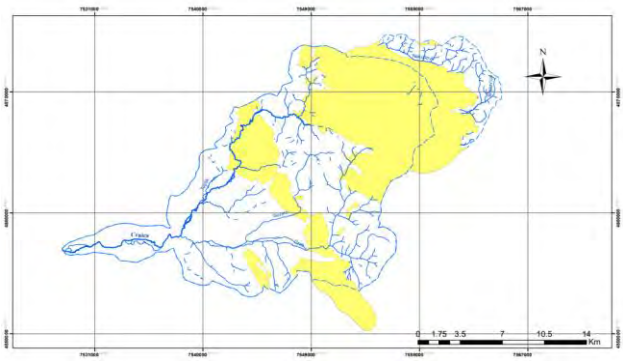


Figure 1. Hydrographic map of the Crnica River, with the karst area indicated

2. METHODOLOGY

For the purposes of preparing this paper, all available data on monthly and annual precipitation sums, mean monthly and annual air temperatures, as well as mean monthly and annual discharges of the Crnica River were collected. The total observed period available was from 1949 up to and including 2022 (meteorological data), and from 1926–2022 (hydrological data). Since the observed discharge consists of direct runoff, which is the consequence of precipitation, and base runoff, which is formed through the recharge of surface flow by groundwater, direct runoff was first separated from base runoff. For this purpose, the software BFI+ (Hydro Office, 2020) was applied, using the Local Minimum method (Wahl & Wahl, 1995). This method identifies days with the lowest flow within a defined time interval — half of the observed period reduced by one day before and after the analyzed day. If the flow on that day is the lowest, it is marked as a local minimum and connected by straight lines with neighboring minima. The base flow between these points is estimated by linear interpolation, whereby the lowest values on the hydrograph are connected into a simplified curve (Sloto & Crouse, 1996).

For monitoring climate change, the World Meteorological Organization proposed and adopted a thirty-year reference period, namely 1961–1990. In this regard, for the analysis of the homogeneity of time series by examining mean values, Student's t-test (for sample size $n < 30$) and the Normalized Z-test (for sample size $n > 30$) may be used. Since in the specific cases we will have separated time series of size 30: the first sample from 1961–1990 and the second from 1991–2020; both of size 30 ($n = 30$), Student's t-test was employed for the homogeneity analysis, in which the criterion for evaluating mean values is obtained using the following expression (Prohaska, 2004):

$$t = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{n_1 \cdot \sigma_{X_1}^2 + n_2 \cdot \sigma_{X_2}^2}} \quad (1)$$

Where n_1 and n_2 are the sizes of time series 1 and 2, $\bar{x}_1$ and $\bar{x}_2$ are the mean values of time series 1 and 2, and σ_1 and σ_2 are the standard deviations of time series 1 and 2.

The variable t follows Student's distribution with $n_1 + n_2 - 2$ degrees of freedom. In this case, the null hypothesis is accepted, i.e., the series is homogeneous, if:

$$t_{\frac{\alpha}{2}} < t < t_{1-\frac{\alpha}{2}} \quad (2)$$

Where α is the significance threshold, and in the specific case the value $\alpha = 0.05$ was taken.

For testing homogeneity through variances, Fisher's F-test was applied. The criterion for the equality of two dispersions is the F statistic, which is calculated using the following expression (Prohaska, 2004):

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad (3)$$

with the condition that $\sigma_1^2 \geq \sigma_2^2$. If the opposite holds, then the F statistic is calculated as:

$$F = \frac{\sigma_2^2}{\sigma_1^2} \quad (4)$$

Thus, the F statistic, which follows Fisher's distribution with degrees of freedom calculated as $n_1 - 1$ and $n_2 - 1$, always has a value greater than 1. The null hypothesis is accepted if:

$$F < F_{1-\alpha}(\nu_1, \nu_2). \quad (5)$$

3. RESULTS

Recharge of the existing aquifers, primarily karst ones, is to the greatest extent carried out through infiltration of atmospheric precipitation, and for that reason emphasis was placed on the analysis of the pluviographic regime as well as the temperature regime (due to the presence of snow cover) characteristic of the catchment area of the Crnica River. For this purpose, precipitation and air temperature values observed at the meteorological station Ćuprija for the observation period 1949–2022 were taken.

Based on the available data on daily precipitation sums observed at the meteorological station Ćuprija, it was determined that the average multi-year precipitation for the observation period 1949–2022 amounts to 666.2 mm. The year with the highest precipitation sum was 2014, with a value of 961.5 mm, while the lowest annual precipitation sum was recorded in 2011, amounting to 432.6 mm. Based on the annual precipitation sums for the observed period, it can be noted that the trend line indicates a slight increase in precipitation amounts (Figure 2).

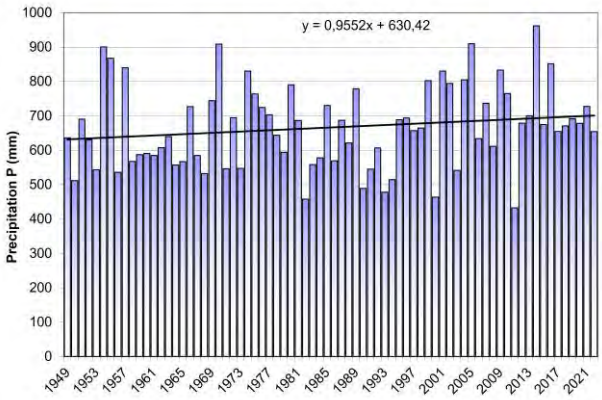


Figure 2. Annual precipitation sums with trend line, meteorological station Ćuprija, period 1949–2022

The graph of intra-annual precipitation distribution (Figure 3) indicates that the rainiest month is June, with an average value of 78.3 mm, followed by May, during which an average of 77 mm falls. The driest months are February (44.3 mm), March (44.9 mm), and August (45.5 mm). In addition, average monthly precipitation amounts below 50 mm are also recorded in January (48.5 mm) and October (47.3 mm). During the observed period, the maximum monthly precipitation sum was recorded in June 1969, amounting to 226.5 mm, while August 2012 stands out as a month without recorded precipitation (0 mm).

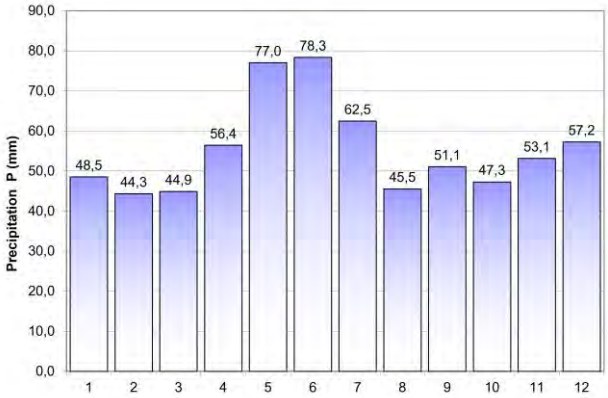


Figure 3. Intra-annual precipitation distribution, meteorological station Ćuprija, period 1949–2022

With regard to mean seasonal precipitation sums, Table 1 presents their basic statistical parameters (mean value, standard deviation, coefficient of variation and skewness, as well as recorded extreme values). In Table 2, the equations of the trend lines for each individual season are given, and based on these equations, the values of expected precipitation sums related to the time horizons 2050 and 2100 were calculated.

Table 1 Statistical parameters of mean seasonal precipitation sums, meteorological station Ćuprija, period 1949–2022

	Winter	Spring	Summer	Autumn
P_{av}	137.7	211.8	159.1	157.7
St.dev.	49.1	71.9	69.5	53.1
Cv	0.4	0.3	0.4	0.3
Cs	0.4	0.8	1.2	0.7
Max	258.0	417.9	400.3	322.0
Min	50.9	77.2	55.2	60.2

Table 2 Trend line equations of analyzed precipitation sums (annual and seasonal) and calculation of their expected values for the time horizons 2050 and 2100

Period	Trend equation	Time horizons		Difference over 100 years
		2050	2100	
Annual	P = 0.9552x+1230.3	727.8	775.6	95.5
Spring	P = -0.0136x+238.7	210.8	210.1	-1.4
Summer	P = 0.0584x+43.129	162.8	165.8	+5.8
Autumn	P = 0.2667x-371.85	174.9	188.2	+26.7
Winter	P = 0.6437x-1140.3	179.3	211.4	+64.4

Based on the results presented in Table 2, it can be concluded that the precipitation trend is generally positive (except in spring, when a distinctly slight negative trend is obtained). The most pronounced increase is expected during the winter months (around 65 mm) and autumn months (around 27 mm on a 100-year scale).

For analyzing air temperature in the Crnica River catchment, mean monthly and annual air temperatures observed at the meteorological station Ćuprija (123 m a.s.l.) during the period 1949–2022 were analyzed. The average multi-year air temperature in the observed period amounted to 11.3 °C. The maximum mean annual value (13.3 °C) was recorded in 2019, while the lowest mean annual value (9.8 °C) was recorded in 1976. In general, the lowest air temperatures occur in the winter months, and the highest during the summer months (July and August). The maximum mean monthly value (25.7 °C) was recorded in August 1952, while the minimum (–7.7 °C) was recorded in February 1956.

In Figure 4, the mean annual air temperatures are presented, together with the trend line, which clearly shows a positive trend of mean annual air temperatures, indicating that during the observed period there was an increase in air temperature of about 1.2 °C. It should also be noted that the increase in air temperature is more pronounced in the last 20 years. More precisely, since the year 2000, only 2005 had a mean annual air temperature lower than the multi-year average, while the others had either the same value (2004, 2006, and 2011) or higher.

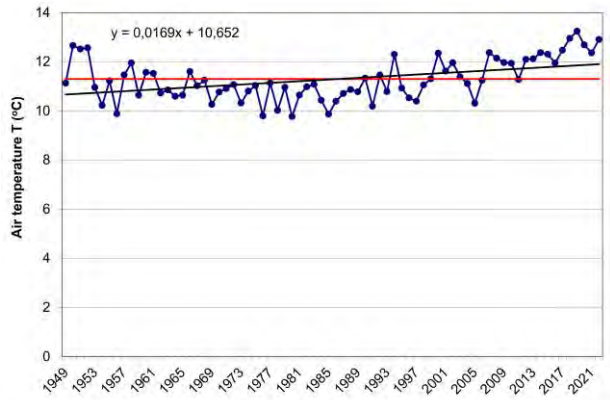


Figure 4. Mean annual air temperatures, meteorological station Ćuprija, observation period 1949–2022

On average, the warmest month is July (21.6 °C), while the coldest is January (0.0 °C) (Figure 5). Regarding the seasons, the warmest is summer (19.9 °C), while the coldest is winter (2.7 °C).

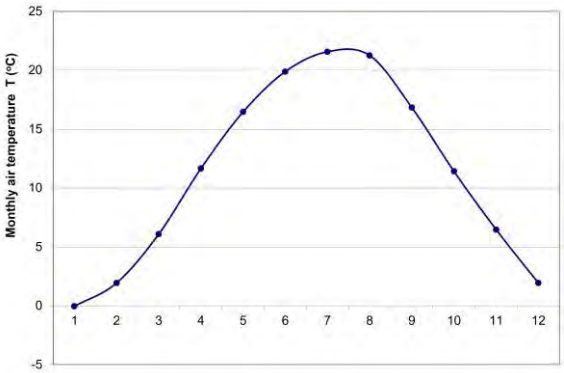


FIgure 5. Intra-annual distribution of air temperature, meteorological station Ćuprija, period 1949–2022

By the same principle as for precipitation sums, Table 3 presents the statistical parameters as well as the trend line equations of mean annual and seasonal air temperatures (Table 4). From Table 4 it can be concluded that the trends are positive in all cases, with the most pronounced trend related to the winter months (the expected increase by the end of this century is 2.7 °C), while the smallest expected increase in temperature during the autumn months (only 0.4 °C).

Table 3 Statistical parameters of mean seasonal air temperatures, meteorological station Ćuprija, period 1949–2022

	Winter	Spring	Summer	Autumn
T_{av}	2.7	16.0	19.9	4.0
St. dev.	1.8	1.0	1.3	0.8
Cv	0.7	0.1	0.1	0.2
Cs	-0.4	0.3	0.4	0.2
Max	6.4	19.1	23.1	6.0
Min	-2.2	13.8	17.3	2.2

Table 4 Trend line equations of analyzed air temperatures (annual and seasonal) and calculation of expected values for the time horizons 2050 and 2100

Period	Trend equation	Time horizons		Difference over 100 years
		2050	2100	
Annual	T = 0.0169x-22.291	12.4	13.2	1.7
Spring	T = 0.015x-13.708	17.0	17.8	1.5
Summer	T = 0.0192x-18.266	21.1	22.1	1.9
Autumn	T = 0.0039x-3.7711	4.2	4.4	0.4
Winter	T = 0.027x-50.907	4.4	5.8	2.7

For the application of homogeneity tests, mean values of precipitation and air temperature as well as standard deviations were first calculated for both separated periods (I: 1961–1990 and II: 1991–2020) — see Tables 5 and 6.

Table 5 Statistical parameters (mean value and standard deviation) of annual and seasonal precipitation sums for the considered separated periods

	Period I		Period II	
	P_{av}	□	P_{av}	□
Annual	648.3	106.8	685.7	129.9
Spring	218.7	76.7	213.2	71.2
Summer	151.5	65,3	160.4	63.9
Autumn	144.9	49.7	165.0	50.2
Winter	133.1	47.9	147.2	63.5

Table 6 Statistical parameters (mean value and standard deviation) of annual and seasonal air temperatures for the separated periods considered

	Period I		Period II	
	T_{av}	□	T_{av}	□
Annual	10.7	0.5	11.7	0.8
Spring	15.5	0.9	16.4	1.0
Summer	19.0	1.0	20.5	1.1
Autumn	3.7	0.6	4.1	0.8
Winter	2.3	1.7	3.1	1.7

Based on expression (1), the values of the criterion t for Student's t-test were calculated, whose numerical values are given in Table 7 for the mean multi-year value and mean seasonal values for the separate periods of precipitation and air temperature. For defining whether the series is homogeneous or not (expression (2)), the critical limits $t_{\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$ were determined, which in the specific case, for the significance threshold $\alpha = 0.05$ and the number of degrees of freedom $\nu = 58$, amount to $t_{\frac{\alpha}{2}} = -2.00$ and $t_{1-\frac{\alpha}{2}} = 2.00$.

Table 7 Values of the t statistic calculated for the purpose of adopting homogeneity of the analyzed precipitation and air temperature series

	Precipitation	Air temperature
Annual	-1.198	-5.708
Spring	0.283	-3.602
Summer	-0.524	-5.434
Autumn	-1.532	-2.154
Winter	-1.056	-1.792

Note: Grey-shaded fields indicate series that are not homogeneous

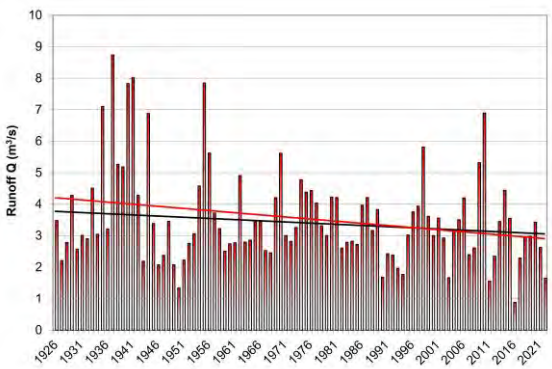
Furthermore, based on expressions (3) and (4), the values of the criterion F for Fisher's F-test were calculated, whose numerical values are given in Table 8 for the mean multi-year value and mean seasonal values for the separate periods of precipitation and air temperature. For defining whether the series is homogeneous or not (expression (5)), the critical limit $F_{1-\alpha}(\nu_1,\nu_2)$ was determined, where the numbers of degrees of freedom are $\nu_1 = 29$ and $\nu_2 = 29$ which in the specific case, for the significance threshold $\alpha = 0.05$, amounts $F_{1-\alpha}(29,29) = 1.153$.

Table 8 Values of the F statistic calculated for the purpose of adopting homogeneity of the analyzed precipitation and air temperature series

	Precipitation	Air temperature
Annual	1.479	2.56
Spring	1.160	1.234
Summer	1.044	1.210
Autumn	1.022	1.777
Winter	1.252	1.000

Note: Grey-shaded fields indicate series that are not homogeneous.

The hydrological station Paraćin, established on the Crnica River, began operation in 1925 when a staff gauge was installed. Later, in 1959, a limnigraph was set up for continuous monitoring of water levels. The analysis of Crnica River discharges characteristic of the Paraćin hydrological station was carried out for the period 1926–2022. The graphical representation of mean annual discharges with trend lines is given in Figure 6.



Fugure 6. Mean annual discharges of the Crnica River observed at the profile of the Paraćin hydrological station with trend lines: red line for the period 1926–2022, black line for the period 1949–2022

Based on the analyzed data for the observation period, the mean multi-year discharge of the Crnica River amounts to 3.56 m³/s. On an annual scale, discharges ranged from 8.749 m³/s (1937) to 0.89 m³/s (2016). The discharge trend line for the entire observation period (1926–2022) indicates a decrease in discharge (Figure 6, red line). The same figure also shows the trend line for the period 1949–2022, i.e., the period for which meteorological data were available (black line). This trend line is also negative, indicating that the discharge of the Crnica River tends to decrease. Since the precipitation trend line is positive, this negative discharge trend may be a consequence of the evident increase in air temperature, particularly pronounced during the summer months, resulting in increased evaporation, as well as increased water demand for irrigation of agricultural land present in the lower part of the Crnica River basin.

Considering the intra-annual distribution (Figure 7), the months with the highest discharges are March and April (slightly below 7.4 m³/s), while September and October are the months when the Crnica River at the Paraćin profile averages only 1.25 m³/s. On a monthly scale, the lowest measured discharge was in November 1990 and 2021, amounting to only 40 l/s, while May 2014 was the month with the highest discharge, amounting to 30.20 m³/s, as well as April 1935, when a discharge of 30.14 m³/s was recorded.

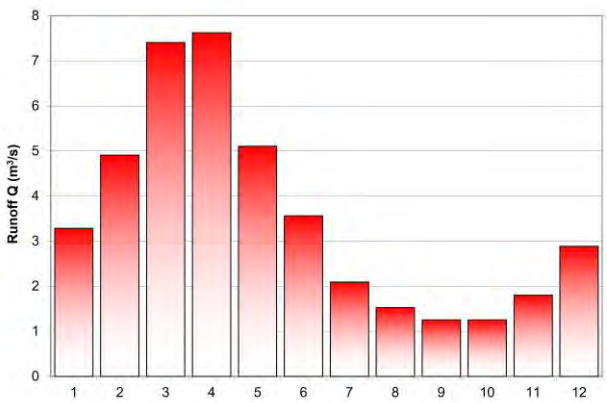


Figure 7. Mean monthly discharges of the Crnica River observed at the Paraćin hydrological station profile, observation period 1926–2022

Based on the observation of discharges by seasons (Figure 8), it can be stated that discharge is disproportionately distributed. In the first part of the year, during winter (37 %) and spring (38 %), about 75 % of the total water volume flows. Summer is the driest season (11 %), followed by autumn (14 %); together, during these two seasons, about 25 % flows, which is approximately three times less compared to the spring and winter periods. The percentage of water volume flowing by seasons indicates that in the non-vegetation period 50.5 % flows, while in the vegetation period 49.5 % flows.

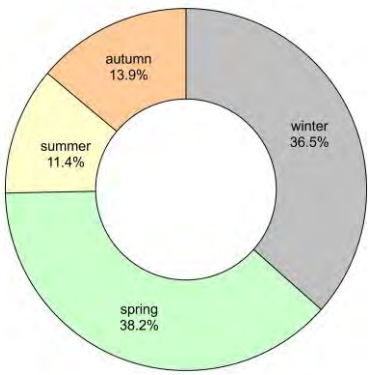


Figure 8. Percentage share of water volumes of the Crnica River separated by seasons

As with precipitation and air temperature, Table 9 presents the statistical parameters as well as the trend line equations of mean annual and seasonal discharges of the Crnica River (Table 10). Based on the trend line equations, it can be concluded that the trends of annual and seasonal discharges are negative. For example, during spring the most pronounced negative trend is observed, i.e., it is expected that by the end of this century the Crnica River will carry even less than 1.74 m³/s during this season. A similar trend is present in autumn, with the expected reduction in discharge during the 21st century being about 1.27 m³/s.

Table 9 Statistical parameters of mean seasonal discharges of the Crnica River, Paraćin hydrological station, period 1926–2022

	Winter	Spring	Summer	Autumn
Q_{av}	5.20	5.43	1.62	1.98
St.dev.	2.51	3.06	1.62	1.52
Cv	0.48	0.56	1.00	0.77
Cs	1.21	1.01	2.86	1.76
Max	14.97	15.80	9.29	7.94
Min	1.48	1.13	0.21	0.20

Table 10 Trend line equations of mean annual and seasonal discharges and calculation of their expected values for the time horizons 2050 and 2100

Period	Trend equation	Time horizons		Difference over 100 years
		2050	2100	
Annual	P = -0.0135x+30.22	2.55	1.87	-1.35
Spring	P = -0.0174x+39.72	4.05	3.18	-1.74
Summer	P = -0.0158x+32.80	0.41	0(-0.4)	-0.41
Autumn	P = -0.0127x+26.96	0.93	0.29	-1.27
Winter	P = -0.0083x+21.62	4.61	4.19	-0.83

For the application of homogeneity tests of discharges, mean values and standard deviations were first calculated for both separated periods (I: 1961–1990 and II: 1991–2020) — see Table 11.

Table 11 Statistical parameters (mean value and standard deviation) of annual and seasonal discharges of the Crnica River for the considered separated periods

	Period I		Period II	
	Q_{av}	□	Q_{av}	□
Annual	3.48	0.89	3.20	3.16
Spring	5.65	2.53	4.92	3.03
Summer	1.58	1.03	1.13	0.80
Autumn	1.78	1.33	1.64	1.00
Winter	4.90	1.82	5.10	3.16

Furthermore, by the same principle as in the homogeneity analysis, the t and F statistics were calculated using expressions (1–5), and their results are presented in Table 12.

Table 12 Values of the t and F statistics calculated for the purpose of adopting homogeneity of the analyzed annual and seasonal discharges of the Crnica River for the considered separated periods

	Statistic *t*	Statistic *F*
Annual	0.972	2.036
Spring	0.996	1.434
Summer	1.858	1.658
Autumn	0.453	1.769
Winter	-0.295	3.015

Note: Grey-shaded fields indicate series that are not homogeneous

Finally, following the method described in the section devoted to methodology, baseflow and direct runoff separation was performed. The calculation period for which daily values of total discharge were available is 1961–2022, and for this period values of direct and base runoff were calculated. Accordingly, Figure 9 shows the representation of total and calculated baseflow for the period 1961–1970. The entire period 1961–2020 is not presented graphically due to the length of the time series and the impossibility of displaying all data in a single diagram. Based on the results obtained, Table 13 presents the statistical parameters of the calculated mean annual values of direct and base runoff, with calculated statistics related to the separated periods (I and II).

Table 13 Statistical parameters of mean annual values of direct and base runoff of the Crnica River calculated for the selected time periods: I (1961–1990) and II (1991–2020)

	I period		II period	
	Qd	*Qb*	*Qd*	*Qb*
Mean	1.405	2.056	1.365	1.839
□	0.532	0.574	0.866	0.669
Cv	0.379	0.279	0.635	0.363
t	0.212	1.326		
F	2.650	1.358		

Note: Grey-shaded fields indicate series that are not homogeneous

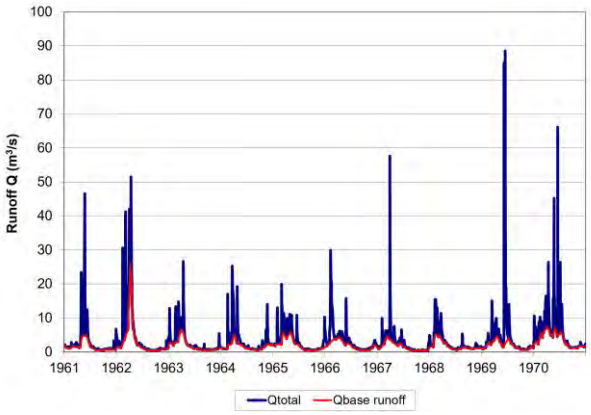


Figure 9. Hydrograph of total and base runoff, Crnica River

4. DISCUSION WITH CONCLUSION

Based on the calculated mean monthly and annual precipitation sums and air temperatures, and the applied trend and homogeneity tests, it has been confirmed that by the end of this century mean precipitation will not change significantly; however, extreme events will become more pronounced and frequent, as indicated by the homogeneity test of variance. Trend analysis showed that on the annual scale the expected increase in precipitation will be only 9.5 mm, i.e., over 100 years this value amounts to 95.5 mm. When this value is distributed by seasons, it can be concluded that precipitation will increase by 2.7 mm during autumn, 6.4 mm during winter, while during summer and spring there will be almost no change. Thus, based on the analyses obtained, it can be stated that no significant changes in mean multi-year precipitation should be expected at either the annual or seasonal level.

Regarding air temperature, homogeneity tests indicate that a statistically significant change in the homogeneity of temperature series has occurred. Trend analysis showed that the expected 100-year increase in temperature will be about 1.7 °C. A similar situation is observed in spring and summer; more precisely, increases of 1.5 °C (spring) and 1.9 °C (summer) are expected. During autumn, no significant increase in air temperature is expected (0.4 °C); however, during winter months the increase may reach almost 3 °C (2.7 °C). Thus, it can be stated that air temperature will increase both annually and seasonally, with the most pronounced increases occurring in summer and winter. These changes, indicating further warming, will lead to increasingly frequent absence of snow cover. Since recharge of aquifers is largely dependent on snowmelt and spring precipitation, due to rising temperatures and the general absence of snow precipitation, this component will have a diminishing effect on groundwater recharge. As more than 75 % of the population in Serbia is supplied with water from groundwater, the population will inevitably face this problem soon.

In addition to mean monthly and annual precipitation, seasonal values of mean discharges were calculated by the same principle. In general, the discharge trend by seasons is negative, with the steepest slope of the trend observed for the summer season, while the mildest decrease can be expected during the winter months. More precisely, if the obtained trend line equations are applied for the entire period and for each season individually, it can be stated that on the annual scale discharge will decrease by about 42 %. During winter months the expected decrease will be only 16.5 %, during spring 35 %, during autumn 81 %, while during summer it amounts to 132 %. By applying the trend equation, it is obtained that the expected discharge during summer will be –0.38 m³/s in the year 2100, which means that the Crnica River will dry up during summer months by the end of this century.

It should be noted that trend analysis generally depends on the selected period itself. The overall trend for the entire observation period has already been presented. However, if trend analyses of mean annual discharges are performed for selected periods (1931–1960, 1961–1990, 1991–2020), a completely different picture is obtained for individual periods (Figure 10). For example, the trend analysis of Period I (1931–1960) shows that during this period there was an exceptionally negative trend. More precisely, if someone had conducted a trend analysis of this period in 1961, they would have concluded that the Crnica River would begin to dry up during summer months by 2016. We are witnesses to say that this was not the case. Furthermore, the discharge trend in Period II is still negative but with a much milder decrease, which is far more acceptable compared to the trend of Period I. Finally, the trend of Period III indicates that in the future an increase in discharge can be expected, i.e., completely opposite to the trends of Periods I and II and the overall period.

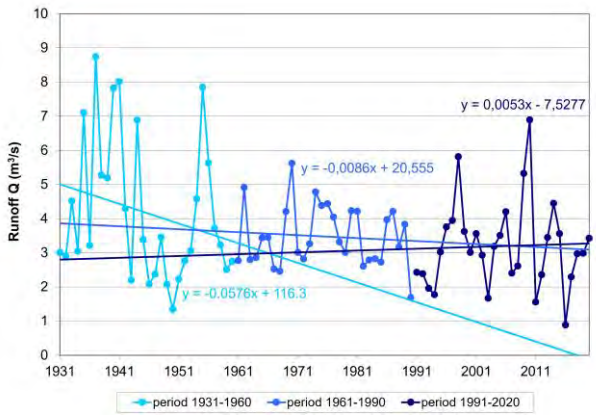


Figure 10. Trend analyses of mean annual discharges calculated for separated time series

The conducted homogeneity analyses of mean values of the separated discharge time series indicate that no statistically significant changes occurred, i.e., the test showed that the time series are homogeneous.

By applying the homogeneity test of variances of the examined samples of annual and seasonal mean discharges of the Crnica River calculated for the considered periods, it is obtained that all analyzed time series are not homogeneous. Such results are certainly influenced by the fact that the presence of extreme events is increasingly pronounced in the case of precipitation and consequently in discharge.

Finally, trend analysis of direct and base runoff was performed, as well as homogeneity testing of these time series. Accordingly, Figure 11 presents the diagram of changes in mean annual values of direct runoff, while Figure 12 presents base runoff. From the diagram of changes in direct runoff (Figure 11), it can be stated that during the analyzed period no significant change in the homogeneity of this time series occurred on the one hand, while trend analysis indicates that no significant changes are expected regarding the mean values of direct runoff. However, the homogeneity test performed using the F-test indicates that the time series is not homogeneous, i.e., it shows that increasingly pronounced extreme events are present.

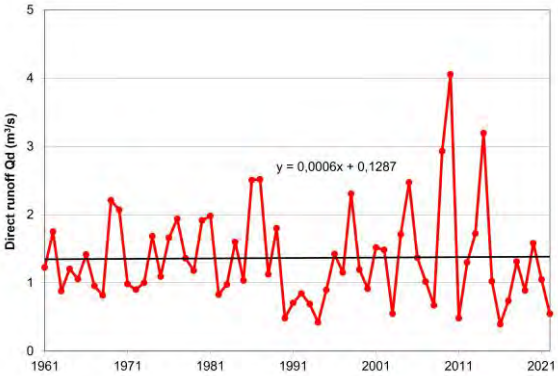


Figure 11. Diagram of changes in mean annual values of direct runoff of the Crnica River

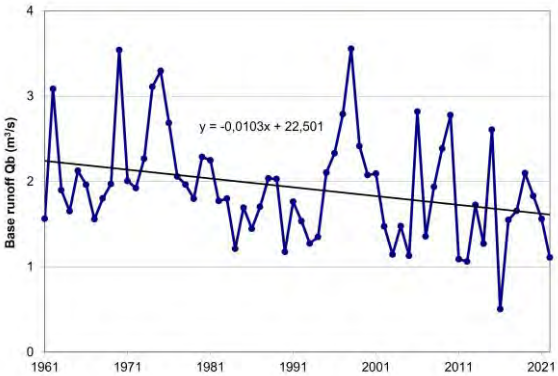


Figure 12. Diagram of changes in mean annual values of base runoff of the Crnica River

Baseflow is important because it is formed from groundwater present in the catchment. The intensity of groundwater recharge decreases over time depending on catchment characteristics, geological structure of the terrain, and hydrogeological conditions, i.e., aquifer types. Trend analysis of mean annual values of baseflow indicates that a decrease of this component of runoff can be expected over a longer period. However, homogeneity tests indicate that the separated time series belong to the same population, whether the t-test or F-test is applied. Thus, the analyzed impact of climate change on meteorological data will not have a significant effect on baseflow. The negative trend is most likely a consequence of the expected increase in temperatures during winter months, when snow cover will be increasingly absent, thereby reducing groundwater recharge. This will affect aquifers poor in dynamic groundwater reserves (fractured aquifers or so-called terrains poor in groundwater). Supporting this is the fact that in 2024 water was delivered by cisterns to livestock grazing on Stara Planina (fractured aquifer present). The reason was that most springs, which had never dried up before, dried up in 2024 due to the absence of snow and snow cover during winter months and a long period without rainfall during summer with extremely high temperatures.

In contrast, karst aquifers have significant porosity and can therefore store considerable dynamic groundwater reserves, which during dry periods are discharged through karst springs over a longer time, so that rivers originating from karst springs baseflow will be more significant, i.e., these streams will not dry up in most cases. In the specific case, the upper part of the Crnica River catchment is predominantly composed of karst terrains (Figure 1).

Finally, it is important to note that if static groundwater reserves exist in karst aquifers (reserves located in the catchment below the elevation of aquifer outflow/discharge), by regulating the karst aquifer these reserves can be used for water supply or other purposes (e.g., irrigation). These reserves will be replenished during rainy periods.

Acknowledgment

Contract on realization and financing of scientific research of SRI in 2026, No.451-03-34/2026-03/200126, and No.451-03-33/2026-03/200126, Ministry of Science, Technological Development and Innovation of Republic of Serbia

References

[1] Hydro Office. (2020). Software for hydrology, meteorology, geoscience & data science. – Available from: https://hydrooffice.org.

[2] Prohaska, S. J. (2004). Hidrologija - I deo. Rudarsko-geološki fakultet, Beograd.

[3] Sloto, R. A., Crouse M. Y. (1996). HYSEP: A computer program for streamflow hydrograph separation and analysis. – U.S. Geological Survey, Water-Resources Investigations Report 96-4040, Pennsylvania, 46 p.

[4] Stevanović Z., Milanović S. (2017). Metode hidrogeološkog istraživanja. Univerzitet u Beogradu, Rudarsko- geološki fakultet, Beograd, str 1-552.

[5] Vakanjac J., Tanasković A., Nikolić J., Ristić Vakanjac, V., Vakanjac, B., Lončar J. (2025), Trend analysis of precipitation and air temperature observed at the meteorological station Belgrade, SymOpIs25, Palić

[6] Vakanjac J., Tanasković A., Nikolić J., Lončar J., Polomčić D., Vakanjac B., Ristić Vakanjac V. (2025), Analysis of the homogenity of precipitation and air temperatures observed at the Belgrade meteorological station, Međunarodna konferencija SETI VII

[7] Wahl, K. L., Wahl T. L. (1995). Determining the flow of Comal Springs at New Braunfels, Texas. – In: Proc. Texas Water ’95, component conference of the First Int. Conf. on Water Resources Engineering. San Antonio, American Society of Civil Engineers, 77–86.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EVALUATING THE EFFECTIVE PERFORMANCE OF THE PHASE NOISE MEASUREMENT SYSTEM

NEDA SPASOJEVIĆ
Technical Test Center, Vojvode Stepe 445, Belgrade, Serbia, nedamilivojcevic@yahoo.com

MIŠA MARKUŠ
Technical Test Center, Vojvode Stepe 445, Belgrade, Serbia, markus.misa@gmail.com

SLOBODAN SUBOTIĆ
Technical Test Center, Vojvode Stepe 445, Belgrade, Serbia, slobodansubotic80@gmail.com

DRAGAN LAZIĆ
Technical Test Center, Vojvode Stepe 445, Belgrade, Serbia, astazu.lazic@gmail.com

Abstract: *The residual phase noise (also known as the noise floor) of a phase noise analyzer directly determines the sensitivity of that standard, as well as the overall performance of the phase noise measurement system. Since the ML 02 laboratory of the Technical Test Center possesses this standard within its frequency stability and phase noise measurement system, the aim of this research was to experimentally determine whether the specified noise floor is achievable, and whether the actual noise floor is lower than the specified values. This provides insight into the real performance of the existing system and the minimum level of phase noise that can be measured using this setup. This article presents the system noise floor results obtained for different measurement times across the entire frequency offset range of the analyzer, as well as the measurement method and conditions under which these results were obtained.*

Keywords: *phase noise, noise floor, residual phase noise, frequency stability.*

1. INTRODUCTION

Phase noise arises from small, random fluctuations or uncertainties in the phase of an electrical signal. The limitation of dynamic range due to the presence of phase noise directly constrains the performance of various systems. In radar systems and telecommunications, this manifests as reduced system sensitivity; in image processing, as diminished resolution; while in digital systems, for instance, it appears as an increased bit error rate [1]. In analog audio communication systems, the impact of unwanted noise is perceived as a “whistling” within the system’s frequency range [2]. Considering the influence of phase noise on the performance of individual systems or their components, it is essential to measure this parameter in order to assess their reliability.

Phase noise in oscillators originates from random noise sources including flicker noise, shot noise, thermal noise etc. These mechanisms dominate over different offset frequency ranges, and their contributions depend on the oscillator architecture and the underlying component technology.

Due to physical implementation, every oscillator or frequency generator can be regarded as a filtered noise source [3]. Consequently, noise exists around the fundamental frequency and represents the equivalent of random amplitude and frequency modulations of an ideal signal.

However, it is important to distinguish continuous (random) phase variations of the signal from discrete ones. Discrete phase changes give rise to discrete signal components, known as spurious, which do not constitute phase noise. Phase noise is caused by continuous phase

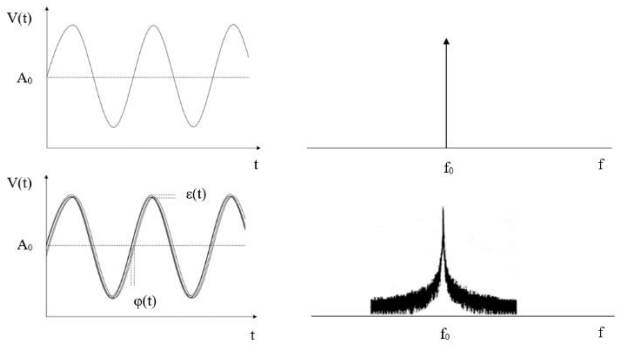


Figure 1. The ideal signal (top) and the real signal in the presence of amplitude and phase fluctuations (bottom), represented in both the time and frequency domains

fluctuations, which in the frequency domain are observed as a broadening of the signal spectrum [4]. Figure 1 illustrate the ideal (top) and real (bottom) electrical signal in both the time and frequency domains.

Phase noise is formally defined as the ratio of the noise

Corresponding author: N. Spasojević;
E-mail: nedamilivojcevic@yahoo.com
DOI: 10.5937/OTEH26148S

power measured within a 1 Hz bandwidth at an offset frequency close to the carrier (P_n) to the total power at the carrier frequency (P_c), according to Equation (1) and Figure 2. In technical specifications, this parameter is commonly expressed as single-sideband (SSB) phase noise. It is most commonly denoted as L(f), in units of dBc/Hz.

$$\mathcal{L}(f) = 10log(\frac{P_n}{P_c}) = P_n\left(dBc/_{Hz}\right) - P_c(dBm) \qquad (1)$$

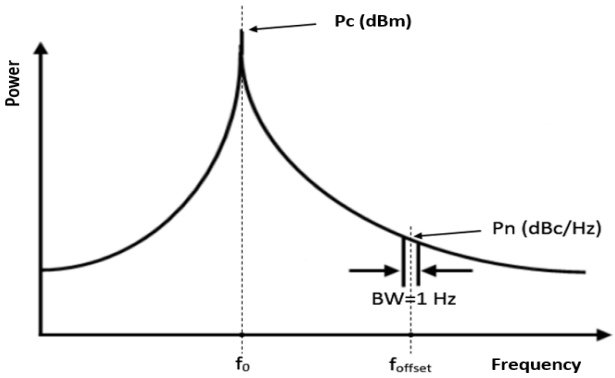


Figure 2. Phase noise, defined as the ratio of the noise power within a 1 Hz bandwidth to the total signal power.

The most common classification of phase noise measurement methods distinguishes six fundamental approaches: the use of a frequency spectrum analyzer, delay line discriminators, quadrature techniques, FM discriminators, cross-correlation methods, and direct digital measurement. These approaches differ depending on whether phase noise is assessed directly, through signal measurement, or indirectly, by calculation from other parameters. They also vary with respect to the number of oscillators employed, specifically whether they require the use of a reference oscillator with known phase noise characteristics [2][5].

At the Technical Test Center, within the Laboratory for Microwave Technics, Time and Frequency (ML 02), a system for measuring phase noise and frequency stability in the time domain has been established. This system employs a phase noise analyzer and an ultra-stable quartz oscillator as a reference standard. In order to assess the actual sensitivity of the system, the analyzer's noise floor as one of its primary limiting factors was measured. In this paper it is also denoted as residual phase noise or internal noise. Through careful selection of measurement method and appropriate configuration settings, the minimum level of phase noise that can be measured with this analyzer was determined.

2. MEASUREMENT METHOD

The implemented phase noise measurement system is based on the use of the phase noise analyzer. The analyzer operates on the principle of digital signal processing, employing a cross-correlation method for signal comparison. In contrast to analog methods, which rely on converting phase fluctuations into proportional voltage levels that are subsequently sampled and processed using Fourier analysis, the digital processing in this system is performed directly on samples obtained from the original test and reference signals, while phase difference detection is carried out digitally.

The phase noise analyzer operates by comparing two signals, while one of them serves as a reference. For this reason, in addition to the analyzer, the measurement system includes an ultraclean quartz oscillator as reference standard. The application of a specific cutting technology of the quartz makes this standard a suitable reference in terms of short-term frequency stability and phase noise. These two standards, each with its own performance characteristics, jointly determine the overall performance of the phase noise measurement system. This system constitutes a unique configuration for direct phase noise measurement within the country. The system for frequency stability and phase noise measurement is configured as shown in Figure 3.

Figure 3. System for frequency stability and phase noise measurement

In order to eliminate the influence of environmental interference, the system is placed inside a modular Faraday cage whose frequency range goes up to 1 GHz and attenuation exceeds 85 dB across the entire frequency range. During measurements, the cage door remains closed, while the environmental conditions are maintained within required limits. The start and stop of the measurement process can be performed either manually or controled via PC and Putty software.

The system noise floor was determined by comparing two identical signals, that represent two copies of a single quartz oscillator output, as shown in Figure 4.

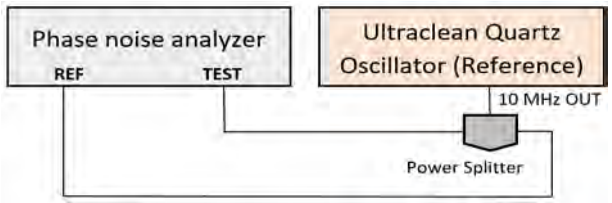


Figure 4. The block diagram of the measurement method

The 10 MHz signal from the output of the quartz oscillator is simultaneously applied to the reference and test inputs of the analyzer through symmetrical power splitter. Although quartz oscillator signals were used for this purpose, in general, the procedure can be performed using any two identical signals.

The analyzer measures the phase noise of input signals with nominal frequency from 2 MHz to 100 MHz, at offset frequency range from 1 mHz to 1 MHz. The signal levels of the test and reference signals must lie within the range of −10 dBm to +20 dBm, while their input levels and nominal frequencies can be different.

The noise floor of the phase noise analyzer, and consequently of the implemented phase noise measurement system, was evaluated for different measurement time: 5 minutes, 15 minutes, and 30 minutes.

During measurements, it is essential to properly configure specific analyzer parameters. The equivalent noise bandwidth (ENBW) should be set to its minimum value of 5 Hz. This increases the capability of measuring very low noise levels and their fine fluctuations. In addition, the thermal compensation and threshold options must be disabled in order to obtain realistic information about the actual internal noise of the analyzer. Enabling the option for thermal compensation suppresses drift and noise caused by temperature variations during the normal operation of analyzer. The threshold option implies rejection of signals below a predefined level in order to separate the analyzer's internal noise from the measured noise of the DUT (Device Under Test). Since the aim of this measurement is to detect all noise components of the analyzer, including its smallest fluctuations, disabling these options ensures that the entire internal noise of the analyzer is involved in the measurement and is not masked.

3. MEASUREMENT RESULTS

The measured residual phase noise of the analyzer for a measurement interval of 15 minutes, presented in the form of a report, is shown in Figure 5. Very low values of residual phase noise can be observed, even at offset frequencies close to the carrier, reaching levels of −191 dBc/Hz within the offset frequency range up to 1 MHz.

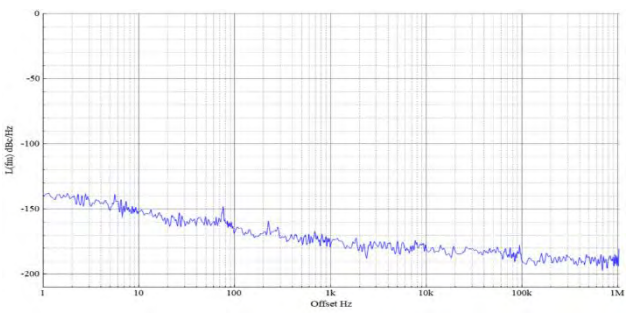


Figure 5. Residual phase noise, 15 min measurement time

Similar results were obtained for measurement intervals of 5 minutes and 60 minutes. The measured values, together with their associated measurement uncertainties and in comparison with the values specified by the manufacturer, are presented in Figure 6. At certain offset frequencies, the measured residual phase noise is up to 7 dB lower than the specified values. The system measurement uncertainty is 2.3 dB, with the dominant contribution arising from the phase noise measurement accuracy of the analyzer. Although the typical value of this accuracy is 1 dB, the manufacturer specifies 2 dB as the guaranteed accuracy of the analyzer [6].

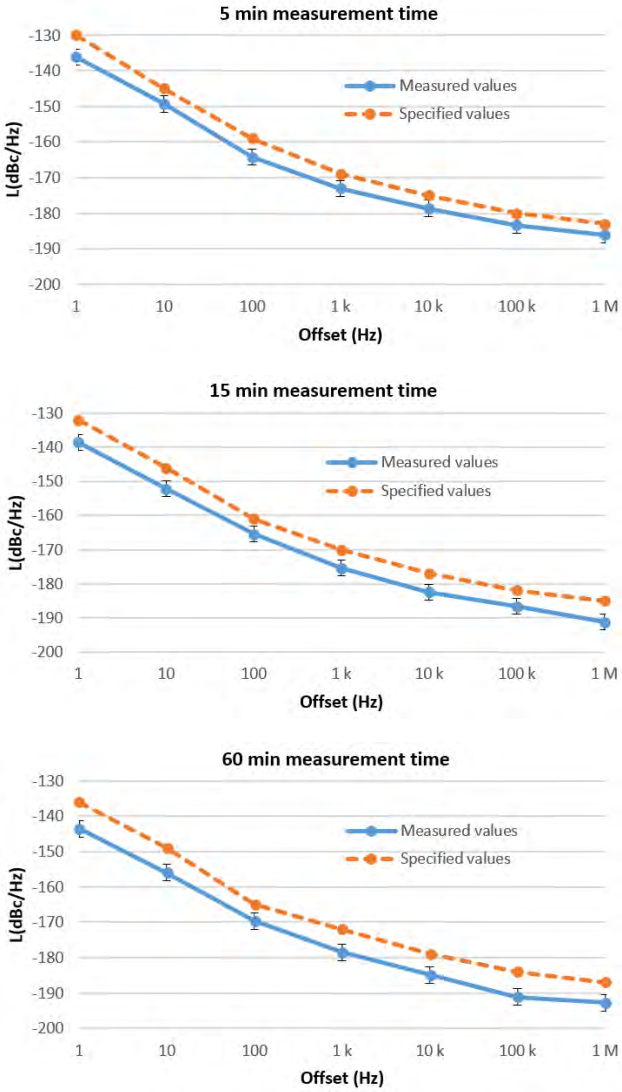


Figure 6. Residual phase noise results in comparison with specifications

Experimental results show that the measured residual phase noise of the analyzer for these measurement intervals (5 min, 15 min, and 60 min) is lower than the specified values, thereby confirming the analyzer specifications. Therefore, such a system enables phase noise measurements of the highest-quality oscillators that are calibrated in the ML02 laboratory.

5. DISCUSSION

Section 3 describes conditions and analyzer settings required to ensure the most realistic measurement of its noise floor (residual phase noise). This includes using of the power splitter, disabling thermal compensation and threshold correction functions, as well as applying the lowest possible equivalent noise bandwidth.

Although these correction functions brings some benefits in the phase noise measurement of the DUT, in the case of internal noise evaluation, they may introduce additional

noise or systematic measurement errors. Consequently, the influence of these factors and their proper settings on the measurement of the analyzer's residual phase noise has been analyzed.

Figure 7 illustrates the difference in residual phase noise when thermal compensation and threshold correction functions are enabled, compared to the case in which these corrections are disabled. Corresponding difference is dominant at higher offset frequencies from 100 kHz and above.

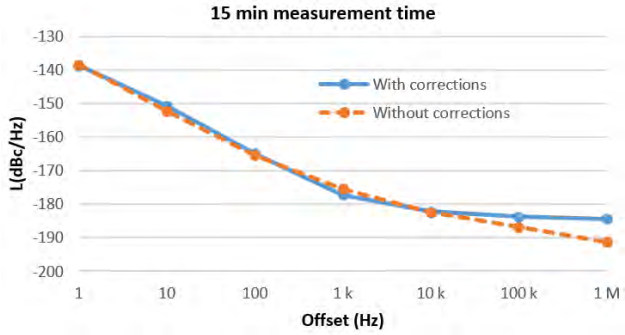


Figure 7. Effects of applying thermal compensation and threshold correction functions on residual noise measurement

Figure 8 presents comparative results of the residual phase noise obtained with the equivalent noise bandwidth (ENBW) set to its maximum value of 500 Hz and its minimum value of 50 Hz. The observed difference is greatest at the 1 MHz offset frequency and amounts to 2.5 dB. Although this parameter does not contribute in obtaining more realistic residual phase noise results in the same manner as the previously described factors, its influence should nevertheless be taken into account in order to achieve the highest possible measurement precision.

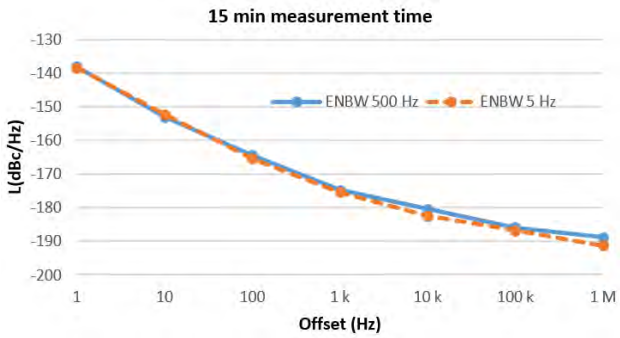


Figure 8. Effects of ENBW on residual phase noise measurement

In Figure 9 is shown comparative results of the analyzer residual phase noise obtained using a power splitter and by applying the 10 MHz signal from two physically separated outputs of the quartz oscillator. Presented results show that the use of a splitter is crucial and necessary for measuring the internal noise of the analyzer. The difference between the presented results is particularly significant at offset frequencies higher than 1 kHz.

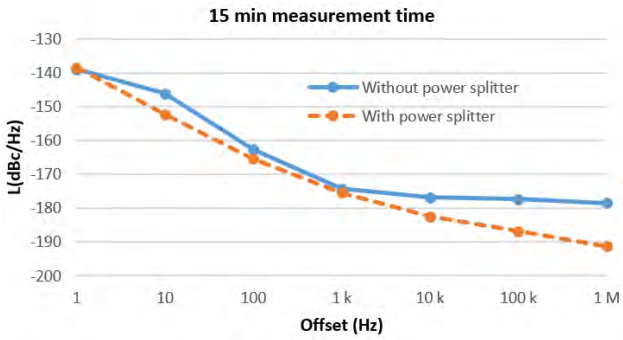


Figure 9. Influence of power splitter on residual phase noise measurement

Considering the finite noise floor of the analyzer, care must be taken to ensure that the phase noise of the oscillator under calibration is not close to this limit. Thanks to the very low noise floor of the applied analyzer, being significantly lower than the phase noise of the oscillators under test, this analyzer fully satisfies the current laboratory requirements for phase noise measurements.

5. CONCLUSION

The paper presents the experimental procedure for determining the residual phase noise of a phase noise analyzer, which defines the noise floor and the effective performance of the phase noise measurement system in the metrological laboratory of the Technical Testing Center. The described method includes the specific measurement conditions and setup that must be met in order to obtain relevant values for the noise floor. The obtained results confirmed the specifications and demonstrated the actual capabilities of both the analyzer and the entire system. The existing system fully meets the measurement requirements for phase noise, serving both the ML 02 laboratory and the entire system of defense.

References

[1] Oscilloquartz, *How to Synchronize Telecommunications Networks,* White paper No.10, Precision, Stability, Innovation, Support

[2] OWEN,D., *Good Practice Guide to Phase Noise Measurement,* Measurement Good Practice Guide No.68, Pickering Interfaces, NPL-National Physical Laboratory, May 2004

[3] RF, RFIC & Microwave Theory, Design, *Phase Noise Basics,* www.rfic.co.uk

[4] ROBERTS,P., *Fluke Precision Measurement Ltd, Understanding Phase Noise in RF and Microwave Calibration Applications*, NCSL International Workshop and Symposium, 2008

[5] NELSON,C., *Phase Noise Measurements*, NIST, 2011

[6] NoiseXT, *DNA Series Digital Phase Noise and Frequency Analyzer*, Datasheet, Arcale Spherea Group, 2024

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

IMPLEMENTATION OF A LOW-COST HP-IB (GPIB, IEEE 488) INTERFACE AND MATLAB GUI APPLICATION FOR HP 3582A SPECTRUM ANALYZER CONTROL

PETAR MIŠLJEN
Military Technical Institute, Belgrade, petarmisljen@gmail.com

MILOŠ PAVIĆ
Military Technical Institute, Belgrade, meelos.pavic@gmail.com

SAŠA GUNDELJ
Military Technical Institute, Belgrade, gundelj@yahoo.com

IGOR GORŠIĆ
Military Technical Institute, Belgrade, gorsic@gmail.com

Abstract: *This paper presents a low-cost implementation of an HP-IB (GPIB, IEEE 488) interface based on the open-source AR488 firmware running on an Arduino Nano microcontroller. HP-IB is a short-range, 8-bit parallel, multi-master digital interface bus used for remote control and data acquisition between laboratory instruments and controllers. In this work, the HP 3582A spectrum analyzer serves as a case study.*

The developed system converts data from the GPIB bus to USB serial communication. A graphical user interface (GUI) application was designed in the MATLAB App Designer environment. The application utilizes HP-IB commands to enable instrument connection, parameter configuration, data acquisition, graphic spectrum display, and export of the measured data in ASCII format.

The proposed solution demonstrates how the operational lifespan of legacy laboratory instruments can be successfully extended while allowing their seamless integration and automation in modern laboratory environments. The obtained results confirm reliable communication, accurate data transmission, and a substantial expansion of the HP 3582A's functionality under contemporary laboratory conditions.

Keywords: *HP-IB, GPIB, IEEE-488, AR488, Arduino Nano, HP 3582A, MATLAB App Designer, data acquisition, measurement automation.*

1. INTRODUCTION

The rapid advancement of modern measurement instrumentation has rendered many legacy laboratory devices obsolete, despite their continued technical reliability and high performance. The HP 3582A, a dual-channel dynamic signal analyzer from the late 1970s, exemplifies such equipment [1]. Based on the Fast Fourier Transform (FFT) algorithm, it offers excellent low-frequency spectral analysis capabilities and remains valuable for applications in audio engineering, vibration analysis, and acoustics [2]. However, the lack of native integration with contemporary computers and automated test systems limits its usability in modern laboratory environments. The primary barrier is the obsolescence of the Hewlett-Packard Interface Bus (HP-IB), later standardized as the General Purpose Interface Bus (GPIB) interface under IEEE 488 standard, which, although robust and widely adopted in the past, is no longer natively supported by current personal computers.

This paper presents a low-cost implementation of HP-IB interface based on the open-source AR488 [3] firmware running on an Arduino Nano microcontroller. The developed system enables bidirectional communication between the HP 3582A spectrum analyzer and a personal computer via USB, converting HP-IB commands and data into serial communication. A user-friendly graphical user interface (GUI) application was developed in MATLAB App Designer to facilitate instrument control, parameter configuration, spectrum visualization, and data export in ASCII format. By demonstrating reliable communication and extended functionality of legacy instrumentation, this work contributes to the preservation and modernization of valuable laboratory equipment at minimal cost, offering a practical solution for educational institutions and research laboratories.

2. GPIB (IEEE 488)

One of the longest-standing standards for instrument

Corresponding author: S. Gundelj;
E-mail: gundelj@yahoo.com
DOI: 10.5937/OTEH26149M

interconnection is the GPIB (IEEE 488) [4], [5]. Originally developed by Hewlett-Packard in the late 1960s under the name HP-IB. The GPIB standard was formalized in 1975 and subsequently updated as IEEE 488.1 (electrical and mechanical specifications) and IEEE 488.2 (command and data formats). GPIB (IEEE 488) is an 8-bit parallel bus with a maximum theoretical data rate of 1 MB/s, supporting up to 15 devices over short distances (typically up to 20 m with proper cabling). Its robust handshake mechanism and multi-master capability made it the de facto standard in automated test and measurement systems for decades [3]. Figure 1 illustrates the standard GPIB (IEEE 488) connector pinout.

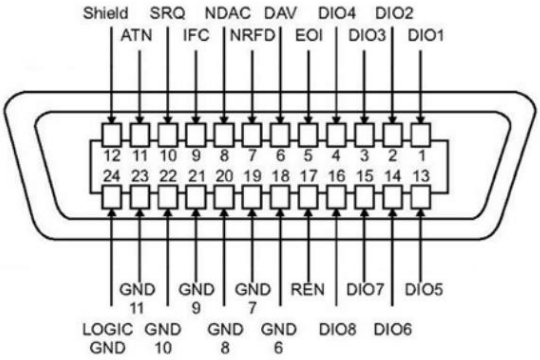


Figure 1. GPIB (IEEE 488) Fimale Connector Pinout [3]

Bus operation is based on dedicated control and handshake lines. The Attention (ATN) line differentiates between command and data modes, while the End Or Identify (EOI) line signals the end of data transfer or enables parallel polling. System control is maintained through Interface Clear (IFC) and Remote Enable (REN) lines, managed exclusively by the system controller. Asynchronous service requests are handled via the Service Request (SRQ) line. Data transfer is synchronized using the three-wire handshake protocol consisting of Data Valid (DAV), Not Ready for Data (NRFD), and Not Data Accepted (NDAC) lines, providing reliable communication even with devices of varying speeds [4], [5]. Figure 2 illustrates the functional grouping of the GPIB (IEEE 488) lines.

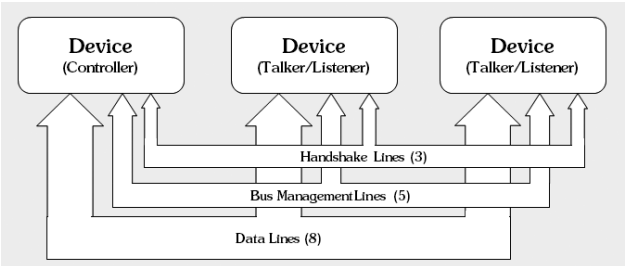


Figure 2. GPIB (IEEE 488) Functional Line Groups [6]

Devices on the bus assume one of three primary roles: Controller, Talker, or Listener. The Controller (usually a computer) manages bus arbitration, addressing, and overall system operation. Multiple controllers are supported, with clear distinction between the System Controller and the active Controller-in-Charge. Addressing is performed using primary addresses from 0 to 30, with Talk and Listen commands derived by adding fixed offsets to the device address. This flexible architecture, supporting configurations from simple point-to-point transfers to complex multi-controller systems, ensures deterministic behavior and remains highly relevant for the modernization of legacy laboratory equipment, as demonstrated in this work.

3. HP 3582A SPECTRUM ANALYZER

The HP 3582A is a classic dual-channel dynamic signal analyzer developed by Hewlett-Packard in the late 1970s. It is based on the Fast Fourier Transform (FFT) algorithm and provides high-resolution spectral analysis in the low-frequency range from 0.02 Hz to 25.5 kHz. Thanks to its advanced digital signal processing capabilities at the time, the instrument achieves a frequency resolution of up to 0.02 Hz with real-time spectrum display, making it particularly suitable for applications in audio engineering, vibration analysis, acoustics, and sub-audio measurements [1], [2]. Figure 3 shows the external appearance of the instrument.

Figure 3. HP 3582A Spectrum Analyzer

Within the HP-IB bus, the HP 3582A offers extensive remote control capabilities. Virtually all front-panel functions are programmable, and the instrument supports full read/write access to its internal memory, including time records, amplitude and phase spectra, transfer functions, coherence functions, and alphanumeric CRT display data [1]. Figure 4 presents a typical display of the analyzer during operation. These features enable efficient integration into automated measurement systems, where data can be systematically acquired, stored, post-processed, and compared with reference values.

From the signal processing perspective, the HP 3582A implements digital heterodyne processing followed by the FFT algorithm, allowing band-selectable analysis with exceptionally high frequency resolution. The instrument supports multiple window functions (rectangular, Hanning, and flat-top) to optimize the trade-off between spectral leakage and amplitude accuracy. Additionally, the built-in sine wave and pseudo-random noise generator (periodic and random) enables standalone testing and calibration of the measurement chain. The dual-channel configuration further supports advanced analyses such as transfer function and coherence measurements — key parameters for system identification and linearity assessment [2]. Due to these characteristics, the HP 3582A serves not only as a measurement instrument but

also as an excellent platform for studying dynamic systems in the frequency domain.

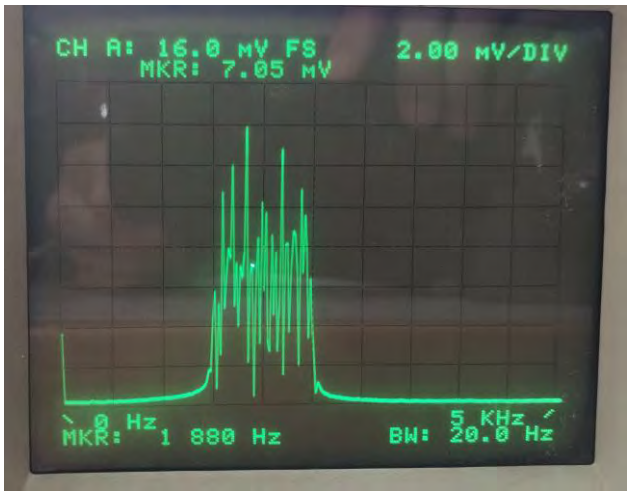


Figure 4. HP 3582A Display Example

4. AR488

The AR488 is an open-source Arduino-based GPIB (IEEE 488) controller developed by Twilight Logic [3], which provides a low-cost alternative to commercial GPIB (IEEE 488) interfaces. The firmware has been developed and tested on various Arduino boards, including the Uno, Nano, Mega 2560, Micro, and Leonardo, with additional support for certain ATmega and ESP32 variants [3]. It emulates Prologix command syntax (with minor exceptions) and enables seamless USB-to-GPIB communication.

The AR488 firmware supports SN75160 and SN75161 GPIB (IEEE 488) transceiver integrated circuits. These ICs provide a buffer between the Arduino and the GPIB (IEEE 488) bus and allow the full 48 mA drive current for a GPIB (IEEE 488) device. In addition, when powered down, these devices present a high impedance to the GPIB (IEEE 488) bus so that the connected device does not interfere with the operation of the bus. [3] In this proposed solution, the SN75160 and SN75161 integrated circuits were not used, as the HP 3582A demonstrated reliable operation in a configuration with one computer. However, for devices requiring higher current drive or full IEEE 488.1 compliance, the use of these IC's is recommended.

Configuration is performed by editing the AR488_Config.h header file, while firmware updates are done through the Arduino IDE. The interface was assembled by soldering short wires from an GPIB (IEEE 488) male connector directly to the Arduino Nano pins, as illustrated in Figure 5. The complete assembly was then placed in a custom 3D-printed enclosure (Figure 6) for mechanical protection and practical use.

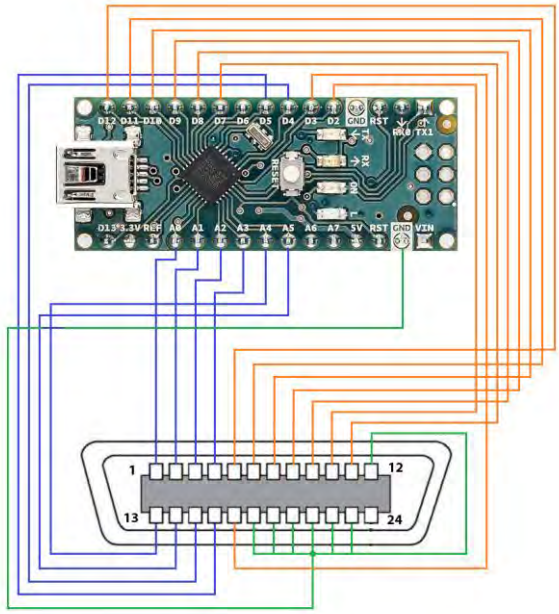


Figure 5. Wiring diagram (male connector)

Testing of the interface was performed using the CoolTerm terminal emulator. The controller version was first verified by issuing the command ++ver, which elicited the response AR488 GPIB controller, ver. 0.53.33, 12/12/2025. Verbose mode was subsequently enabled with the command ++verbose, resulting in the confirmation Verbose: ON.

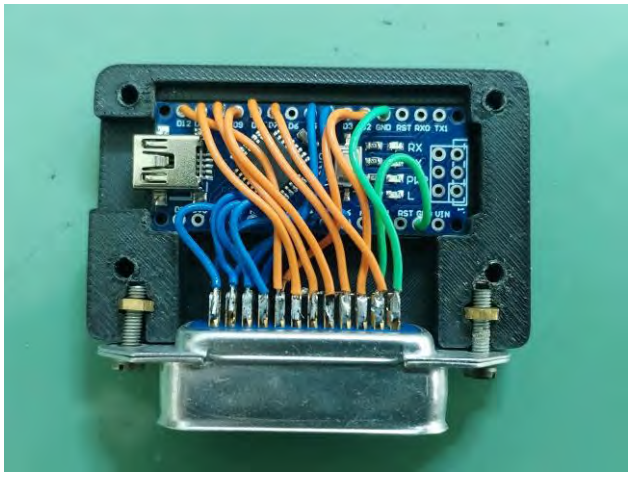


Figure 6. The complete assembly (male connector)

The minimum set of commands required to configure the controller and enable communication with the HP 3582A spectrum analyzer was then applied. The command ++mode 1 configured the AR488 to operate as Controller-in-Charge (CIC) on the HP-IB bus. This was followed by ++auto 1, enabling automatic reading of responses from the instrument, and ++addr 11, which set the HP-IB address of the HP 3582A to its factory default value of 11 (octal). Measured data were subsequently acquired from the spectrum analyzer by issuing the command LDS (Low Data Strobe), which returned a sequence of 256 values. The complete testing procedure is presented in Figure 7.

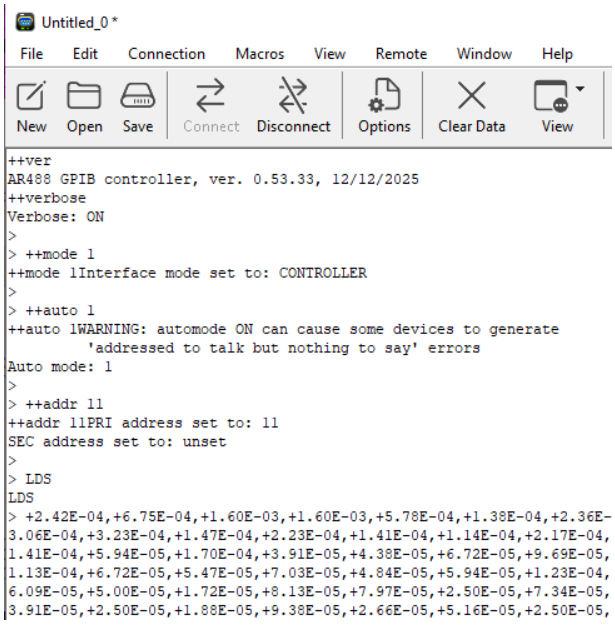


Figure 7. CoolTerm view

This low-cost and compact solution demonstrates the practicality of open-source hardware for extending the usability of legacy GPIB (IEEE 488) instruments.

5. MATLAB GRAPHICAL USER INTERFACE

Prior to developing the full graphical user interface, reliable communication between the AR488 GPIB adapter and the HP 3582A spectrum analyzer was established and tested using MATLAB's serialport interface. The AR488 serves as a USB-to-GPIB bridge, allowing transmission of HP-IB commands over a virtual serial port.

Figure 8 illustrates the initial MATLAB command window test sequence used to configure the adapter and retrieve measurement data.

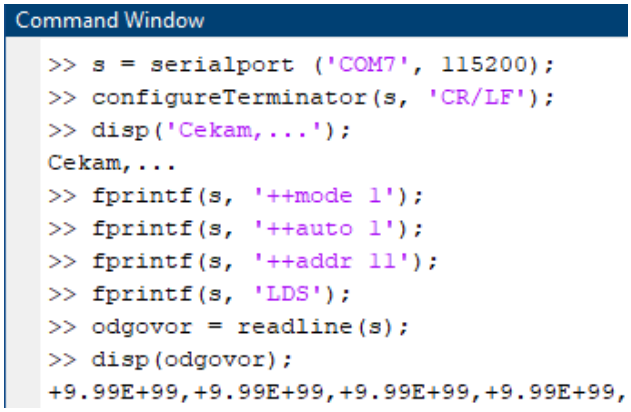


Figure 8. Matlab Command Window

After successful verification of low-level communication, the graphical user interface was designed in MATLAB R2023b App Designer using a modular layout. The main window is divided into three primary sections: (1) a connection and control panel at the top, (2) a central visualization area with axes for spectrum plotting, and (3) a status and data management panel at the bottom. This structure provides an intuitive workflow. Standard App Designer components (drop-down menus, buttons, labels, and axes) were used, with all functionality implemented through callback functions. Figure 9 presents the main window of the developed HP 3582A Control Application (version 1.2).

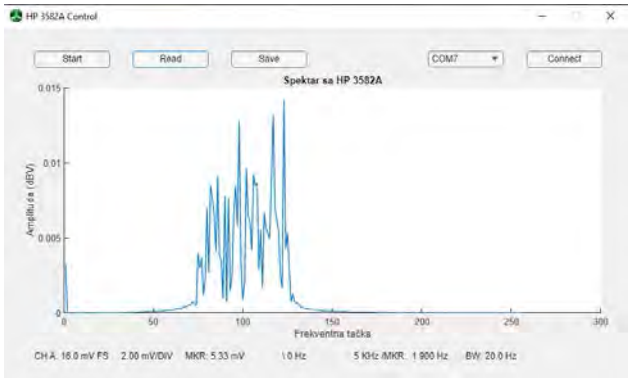


Figure 9. HP 3582A Application V1.2

The core operations are realized as follows. Instrument connection is handled by the ConnectButtonPushed callback function, as illustrated in Figure 10.

```
% Button pushed function: ConnectButton
function ConnectButtonPushed(app, event)
    try
        port = app.COMDropDown.Value;
        app.SerialPort = serialport(port, 115200);  % Baud iz AR488 default
        configureTerminator(app.SerialPort, 'CR/LF');  % ++eos 0 = CR+LF
        app.StatusLabel.Text = 'Čekam inicijalizaciju AR488 (3 sekunde)';
        pause(3.0);
        fprintf(app.SerialPort, '++ver');
        pause(1.0);
        app.StatusLabel.Text = readline(app.SerialPort);
        pause(3.0);
        app.StatusLabel.Text = ['Povezano na ' port '- AR488 inicijalizovan.'];
    catch e
        app.StatusLabel.Text = ['Greška pri povezivanju: ' e.message];
    end
end
```

Figure 10. 'ConnectButtonPushed' function

Upon pressing the Connect button, the application opens the selected COM port at 115200 baud, sets the appropriate terminator, and initializes the AR488 adapter. Comprehensive error handling using try-catch blocks ensures robust operation.

Configuration of the HP 3582A is performed in the StartButtonPushed function, which sequentially sends the necessary commands (++mode 1, ++auto 1, ++addr 11) with appropriate delays (Figure 11).

```
% Button pushed function: StartButton
function StartButtonPushed(app, event)
    try
        app.StatusLabel.Text = 'Setujem HP 3582A (9 sekundi).';
        fprintf(app.SerialPort, '++mode 1');
        app.StatusLabel.Text = 'Setujem Controller mode';
        pause(3.0);
        fprintf(app.SerialPort, '++auto 1');
        app.StatusLabel.Text = 'Setujem auto mode';
        pause(3.0);
        fprintf(app.SerialPort, '++addr 11');
        app.StatusLabel.Text = 'Setujem adresu 11';
        pause(3.0);
        app.StatusLabel.Text = 'Setovanje završeno.';
    catch
        app.StatusLabel.Text = 'Greška pri pokretanju setovanja.';
    end
end
```

Figure 11. 'StartButtonPushed' function

Measurement acquisition and visualization are implemented in the ReadButtonPushed function. When the Read button is activated, the application sends the LDS command, reads the returned data string, converts it into a numeric array using str2double and strsplit, and stores it in the application property SpectrumData. The spectrum is immediately plotted on the designated axes with appropriate labels and title. Additionally, the LAN command is sent afterward to return the instrument to local mode.

Finally, the SaveButtonPushed function enables data export. If spectral data is available, the user is prompted via uiputfile to select a save location. The data is saved as a .txt file in ASCII format with a descriptive header containing timestamp and metadata, ensuring traceability and ease of further processing.

The integration of these callback functions with the graphical components resulted in a robust, user-friendly application that significantly simplifies control of the HP 3582A spectrum analyzer while maintaining full transparency and reproducibility of the measurement process.

During repeated testing sessions, the system exhibited stable operation with consistent spectral data acquisition and no observed transmission errors. The MATLAB GUI provided intuitive and responsive control, noticeably simplifying the measurement workflow compared to manual operation of the HP 3582A.

6. CONCLUSION

This paper successfully demonstrated a low-cost HP-IB (GPIB, IEEE 488) interface based on the open-source AR488 firmware and Arduino Nano microcontroller, coupled with a custom MATLAB GUI application for controlling the HP 3582A spectrum analyzer. The developed system enables reliable bidirectional communication, remote parameter configuration, spectrum acquisition and visualization, as well as data export in ASCII format. By utilizing inexpensive and widely available components, the proposed solution significantly extends the operational life of legacy laboratory instrumentation while seamlessly integrating it into modern computer-controlled measurement environments. The obtained results confirm stable operation and accurate data transmission, validating the practical value of the approach for educational and research purposes.

The presented work opens several avenues for further development. The MATLAB application can be expanded to support virtually all HP-IB commands available on the HP 3582A, enabling more comprehensive automation. Additionally, integration with MATLAB and Simulink environments [7] would allow advanced post-processing of acquired data, such as enhanced filtering, statistical analysis, or system identification techniques not natively supported by the instrument itself. Future research may also explore alternative microcontrollers (e.g., ESP32) to overcome the speed limitations of the Arduino Nano, although the current platform proved fully sufficient for the intended low-frequency applications. Overall, this low-cost modernization strategy contributes to more sustainable and accessible laboratory infrastructure.

References

[1] HEWLETT-PACKARD: *Model 3582A Spectrum Analyzer Operating Manual. Part No. 03582-90002,* Loveland, Colorado, USA, 1981.

[2] HEWLETT-PACKARD: *A High-Resolution, Low-Frequency Spectrum Analyzer.* HP Journal. Palo Alto, California, USA, 1987.

[3] TWILIGHT LOGIC: *AR488 GPIB Controller User Manual*, Twilight Logic, 2025.

[4] THE INSTITUTE OF ELECTRIC AND ELECTRONICS ENGINEERS, INC: *488.1-1987 - IEEE Standard Digital Interface for Programmable Instrumentation*, IEEE, New York, 1988.

[5] THE INSTITUTE OF ELECTRIC AND ELECTRONICS ENGINEERS, INC: *488.2-1987 - IEEE Standard Codes, Formats, Protocols, and Common Commands*, IEEE, New York, 1988.

[6] THE HP 9845 PROJECT, www.hp9845.net, 2026.

[7] MATLAB OFFICIAL WEB PAGE, www.mathworks.com, 2026.

[8] HEWLETT-PACKARD: *HP-IB Command Library, Hewlett-Packard*, Corvallis, Oregon, USA, 1986.

[9] NATIONAL INSTRUMENTS: *GPIB – PC IEEE-488 Instrumentation interface,* National Instruments Corporation, Texas, 1984.

[10] S. AJKALO, O. BJELICA: *One Implementation of IEEE 488 USB Controller,* Elektrotehnički fakultet Istočno Sarajevo, Istočno Sarajevo, 2010.

[11] S. AJKALO, M. GAZIVODA, S. LUBURA, M. ŠOJA: *GPIB and Measurement Instrumentation – Control of Instruments, Acquisition and Presentation of Measurement Data*, Infoteh – Jahorina Vol 6, Istočno Sarajevo, 2007.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

QUASI-DYNAMIC TESTING OF THE QUALITY OF TRACER COMPOSITION IN SMALL ARMS AMMUNITION

BOBAN BOGIĆEVIĆ
Belom d.o.o., bbogicevic@belom.army

NIKOLA KILIBARDA
Belom d.o.o., nkilibarda@belom.army

MIROSLAV PAJČIN
Belom d.o.o., mpajcin@belom.army

OGNJEN PEKOVIĆ
University of Belgrade – Faculty of Mechanical Engineering, opekovic@mas.bg.ac.rs

ALEKSANDAR SIMONOVIĆ
University of Belgrade – Faculty of Mechanical Engineering, asimonovic@mas.bg.ac.rs

Abstract: *This paper presents a quasi-dynamic experimental investigation of the tracer path of a 12.7 × 108 mm projectile, as well as the processing of collected signals using a developed supporting methodology and procedure.*

The aim of the presented methodology and accompanying procedure is to enable reliable, rapid, and safe verification of the quality of the tracer composition within the projectile, using a suitable device, without the need for an open firing range with a long firing line. The experimental procedure described allows for routine and swift factory quality control of tracer projectiles, reducing testing time and cost, improving data acquisition quality and overall test safety compared to conventional field tests, while achieving the desired repeatability of measurement conditions.

The developed methodology, along with the accompanying procedure and device, enables the simulation of real flight conditions of a projectile along its trajectory. The experimental setup allows projectile rotation exceeding 100,000 rpm on the projectile carrier and includes an air-guiding system, a non-contact ignition system for combustion initiation, a spectrophotometer for light emission acquisition, and a PLC for experiment control and monitoring. The results confirm the potential of the developed experimental approach for application in the quality control and development processes of tracer ammunition across various calibers This work relates to the improvements of methods for testing pyrotechnic compositions that make up small arms ammunition tracers, as well as the tracer itself, to the development of contemporary experimental configurations for tracer functional tests, and to an attempt to establish quasi-dynamic testing conditions as an alternative to long-range firing range testing.

Keywords: *small arms ammunition, tracer rounds, tracer testing, pyrotechnic compositions, quality control.*

1. INTRODUCTION AND OBJECTIVES

The manufacturing of cal. 12.7 tracer bullets in the small arms ammunition factory Belom, Požega has been established two years ago. Tracer charge, placed in the container at the bullet's bottom, comprises three different sub-charges, stacked by pressing in sequential order in this container - igniter charge, transfer charge and tracing visible coloured flame producing charge. Corresponding pyrotechnic compositions have been developed and put in production by the factory's resources. Tuning of pyrotechnic compositions, quality control of made batches and lots of loaded tracers include, among the other things, functional tests — emitted colour quality (trace signature quality), burning time, tracing integrity against exposed combined stress in the different phases of bullet firing and motion, and tracer performances at prescribed distances on the bullet trajectory. These tests are principally based on firing tracing rounds, and monitoring and ranking of appearance and colour flame quality by human observers, deployed along the bullet trajectory at assigned distances. Operating procedures and requirements for testing and evaluating of tracer round performances are part of related manuals and standards [1-4]. Tracing signature inspection for every ammunition lot includes shooting 50 to 100 rounds (with double sample repeating in the case of failure), usually at night, and asked for large surface dangerous area because observation posts commonly are distributed in the range at least 1000 m from firing point in the shooting direction and no less than 70 m laterally from the shooting direction. Skilled observers are required for right classification of detected trace signature.

Manufacturing of pyrotechnic compositions, which made tracer charge, is based on the small quantity batches, due to the work safety requirements posed for the high risk

Corresponding author: B. Bogićević
E-mail: bbogicevic@belom.army
DOI: 10.5937/OTEH26150B

processing of pyrotechnic materials. Numerous batches of the same pyrotechnic compositions make corresponding lot. Several lots of the each involved pyrotechnic compositions, i.e. the several tracer lots are necessary for the single tracer ammunition lot. Inspection of each tracer lot are mandatory before assembling in the related ammunition lot. Moreover, pyrotechnic composition lots shall be passed acceptance test before being loaded in to the tracer. Functional testing of tracer is the inherent part of these stepwise quality control process. High volume production of tracer rounds multiplies the amount of this tests, as well as diagnostic of tracer failure mode, if any.

The required long range proving ground is not the part of the Belom factory, therefore scheduling and renting available distant certified proving ground on time represent production organisation and supporting logistic burden.

The existing acceptance test of pyrotechnic lots includes static testing of the burning time and spectral characteristics of the flame emitted by the tracer. Static test method of tracers is proven as not enough efficient in qualifying the effectiveness of the pyrotechnic lots in question and applied process parameters during tracer charge manufacturing. There is real uncertainty about the acceptance tests result of tracer ammunition by shooting, regardless of successful static tracer tests. Complex stress on the tracer charge, and quick changes of combustion conditions during firing and tracer bullet flight (high gradient pressurisation of tracer charge ignition, sharp initial load of inertial force, stress due to imposed high spin, sudden depressurisation on the muzzle exit, vacuum area on the tracer combustion front during bullet flying and shock waves around the bullet due supersonic flow at the tracer acting part of trajectory), can result with the tracer signature or burning time failure due to tracer charge cracking or break-up, de-lamination between tracer sub-charges, and de-bonding from the tracer container.

Acceptance or rejection of tracer lots based on proving ground ballistic tests represents reliable and efficient testing method. However, considering large yearly number of tracer signature testing, testing timeline that corresponding production process, and occasional accessibility to the suitable long range proving ground, have been imposed need for introducing reliable method of testing tracers for burn time prior to proving ground ballistic tests. Moreover, it was attempt to establish alternative acceptance method of tracers in combination with firing tests (EPVAT, precision, function and casualty), which is inherent part of ammunition inspection at the manufacturer site.

There are numerous test fixtures that combined described in the relevant literature [5-7], which combine tracer burn time measurements under simulated bullet's spin as the most influential factor on the tracer burning time. Application of spectrometers for formulation and inspection of the colour flame producing compositions by measuring the spectrum of light from the composition flame is widely used measurement method [8-11], usually applied in a dark environment with evacuation of produced smoke (photometric tunnel).

The skid for testing of tracer signature has been designed by the Belom factory, as an alternative to acceptance and surveillance tests of pyrotechnic compositions that made tracer charge. This measuring skid provides measurement of tracer burn time and emitted spectral characteristics, under the high angular velocities. Additionally, the skid enables evaluation of the possible tracer charge failure mode due to imposed spin. The measuring skid comprises tracer spinning device, ignition source, process controller and spectrometer arranged in to the compact, portable photometric tunnel.

Diagnostic evaluation of tracer failure modes due to rapid pressurisation and depressurisation environment acting on the tracer bullet during barrel motion and at the muzzle exit, is treated as part of regular ballistic testing (EPVAT test, precision test) in the long closed proving range, as well as tracer behaviour during flight up to 100 m from the barrel muzzle. Observations are based on analysing of the video captured by the high speed camera.

The lot of reference tracer bullets has been established by entirely verification of tracer signature and burning time according to the prescribed procedures based on ballistic firing at long range proving ground. After that tracer bullets sample from this reference lot, have been tested by alternative way in order to establish baseline for evaluation of results obtained by testing tracer lot in question. Also, reference lot is using for checking of measuring equipment prior alternative tests are going on.

Combination of these two test methods provide that functioning of the tracer on ballistic flight is properly simulated insofar as is practical and necessary. This complete in-house testing of the tracer functioning represents an alternative to acceptance and surveillance tests of tracer by rounds firing at the long range proving ground.

2. EXPERIMENTAL WORK

The measuring skid (Figure 1) comprises tracer spinner assembly, smoke evacuation fan, optical fibre port, ignition opening and camera for recording, all boxed in to the cube-like container, and external devices attached on the container – laser system for tracer ignition, spectrometer for separation of visible light emitted by tracer to components and measuring them intensities, display for monitoring of ignition and burning process in to container, PLC for ignition and spinning control, electrical box, air compressor, and oil heating and recirculation unit.

Tracer spinner assembly (Figure 2) includes compressed air driven turbine, holders for tracer at turbine rotating shaft, oil lubrication port, and rotational speed indicator.

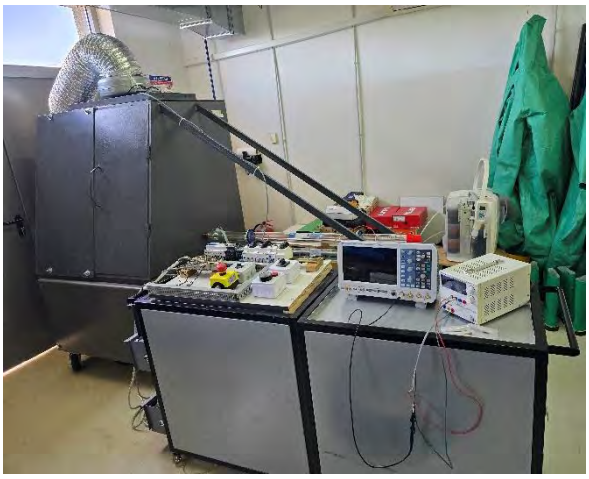


Figure 1. Tracer signature and burning time measuring skid

Turbine has been balanced together with attached tracer up to rotational speed of 200,000 rpm. The rotational speed is adjustable, i.e. applied spin during tracer signature measurement is 128,000 rpm, which corresponding to the rotation rate of 12.7x108 APIT bullet at the barrel muzzle.

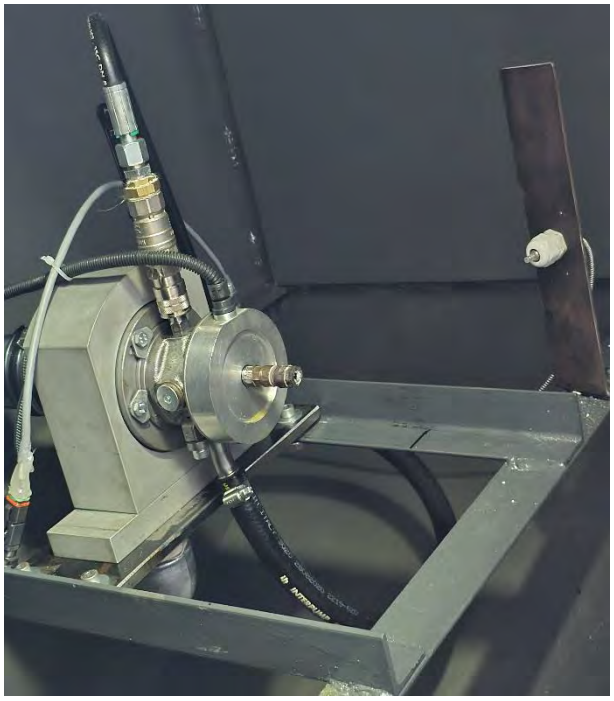


Figure 2. Tracer assembly spinner

The tracer is inserting the core of the rotor and ignited after the desired rotational speed is reached. Ignition has been accomplished by 100W CO2 laser with the laser beam focal spot dimeter of 4 mm (Figure 3).

The laser power is variable, so that the tracer ignition takes place with the minimum energy and exposure time required for stable ignition and combustion of the tracer charge.

The light given off by the burning tracer was picked up by a micro spectrometer with 342 – 820 nm wavelength range, 2048 pixels linear CMOS detector, 16-bit 2 MHz AD converter. The light from tracer signature is transmitted by the optic fibre (Figure 2) to the SMA905 optical input interface on the spectrometer.

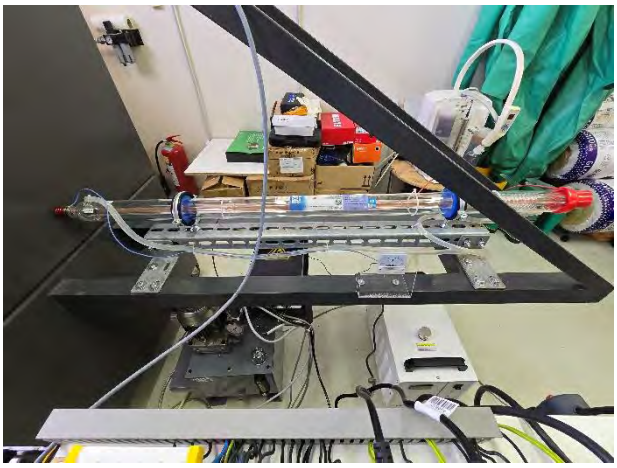


Figure 3. Laser ignition system

When desired laser power is reached tracer charge is ignited and spectral acquisition of emitted light is started. Burning process is monitoring on the external display. Burning time is acquired by the video captured by the camera inside the photometric box. Chronological sequence of tracer burning is given in the Figure 4.

Figure 4. Sequences of tracer burning in photometric box

Acquired spectral data of tracer samples taken for the lot acceptance test, are given in the Figure 5. Calculated tracer colour purity of the red colour is 82.0 to 83.4% (requirement is 75% minimum). Detected burning times are in the range 3.18 – 3.36 s (requirement is 2.6 s minimum).

The test result of pyrotechnic compositions (applied in to the tracer charge) batches, regarding to the spectral characteristics (composition checking, colour purity of tracing composition, burning time measuring) is given in Figure 6.

Observations made from testing of tracer lot under combined stress existing during firing and tracer bullet flight are given in the Figure 7, as series of captured frames, which indicated of proper tracer function.

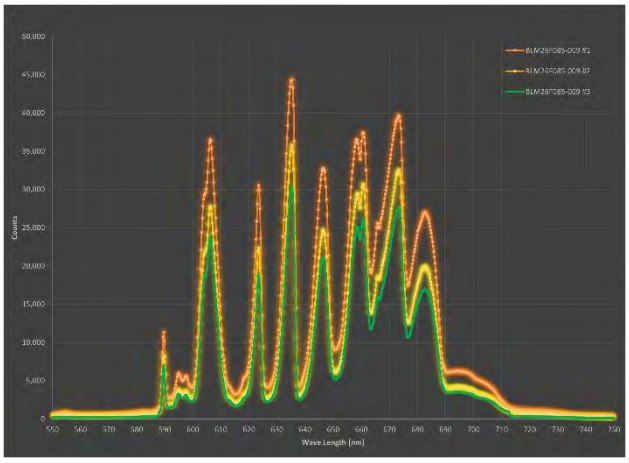


Figure 5. Spectral data for samples from the same tracer lot (acceptance test)

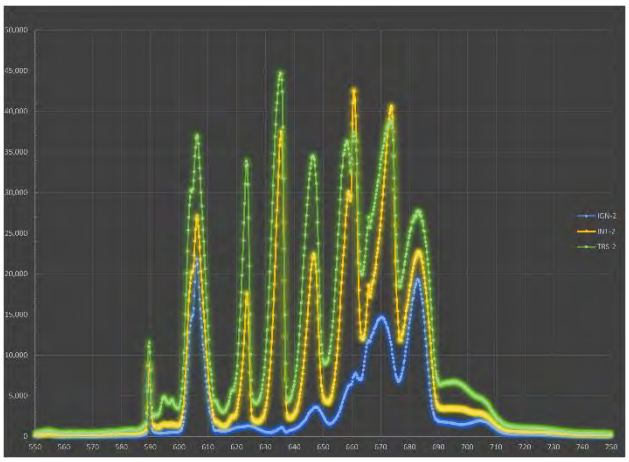


Figure 6. Spectral data for pyro compositions (igniter – blue, transfer – yellow, and tracing – green) which make tracer charge (acceptance test)

Figure 7. Tracer function testing by firing (part of the acceptance test)

The spectrometry offers simply, quick and reliable determination of differences between tracers loaded with sub-charges of different or unknown compositions (Figure 8), with indications of appearance of chemical species in the tracer flame.

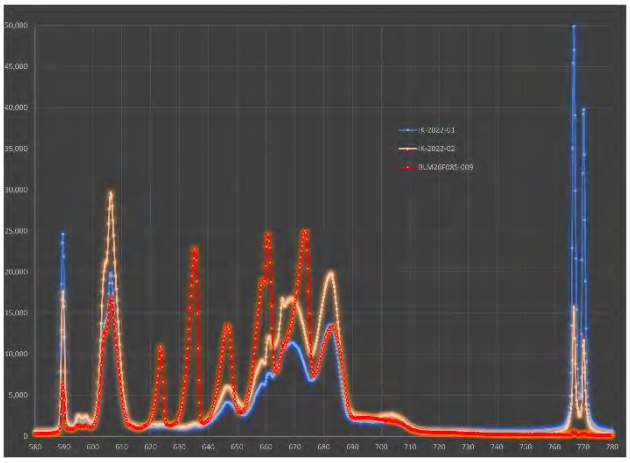


Figure 8. Spectral data for two tracers type regarding to the loaded pyrotechnic compositions (IK and BLM type), but with same tracer configuration (12.7x108 BZT-44 bullet). It is noted absence of SrCl spectra (625, 637, 661, 675 nm), as desired species for red signature in the flame of the IK tracer type, but with strong incandescent emission by the MgO in the flame (760-780 nm), which is undesired because deteriorates color quality of the tracer.

3. CONCLUSIONS

The quasi-dynamic method of tracer signature and burning time measurement has been investigated as an alternative to the firing of tracer bullets at long distances with human observations of the tracer functioning and descriptive evaluation of the flame colour quality.

This method is based on photometric test of tracer flame by the spectrometer, during tracer functioning at spin given by turbine, and corresponding to one imposed by the firing of tracer bullets. Ignition is done remotely by the laser, when assigned spin has been obtained. The related burning time is obtained by the camera. All of these devices are embedded in the portable, compact box.

Firing tracer bullets at short range in-door shooting range supply quasi-dynamic test data with tracer charge structural strength results. It is found that these combined tests represent a real possibility for simple, reliable and timely manner tracer testing at ammunition manufacturer site.

The quasi-dynamic test with ‘reference’ tracers (ones that is tested by the firing procedures at long range proving ground) provides requirements for acceptance test of tracer lot, and for measuring devices calibration and verification.

There is also a useful purpose for this quasi-dynamic test to check the composition and quality of produced batches of pyrotechnic compositions that are included in the composition of the tracer charge.

The spectral data of the tracer signature, obtained by this quasi-dynamic test are valuable in the process of tracer charge formulation, identification and comparative analysis.

References

[1] VOJNOTEHNICKI INSTITUT: *Propisi o kvalitetu proizvoda – Streljačka municija, Metak 12.7 sa zrnima B-32 i BZT-44*, *PKP 1027/84*, VTI KoV, 1984.

[2] NATO STANDARDIZATION OFFICE: *Multi-Calibre Manual of Proof and Inspection (M-C MOPI) for NATO Small Arms Ammunition, AEP-97 Edition A Version 1*, NATO-NSO, 2020.

[3] U.S. ARMY ARMAMENT, MUNITIONS AND CHEMICAL COMMAND: *Small Caliber Ammunition Test Procedures, SCATP-7.62*, U.S AMCCOM, 1990.

[4] U.S. ARMY ARMAMENT RESEARCH, DEVELOPMENT AND ENGINEERING CENTER: *Military Specification, Cartridges, Caliber .50: Armor–Piercing–Incendiary M8; and Armor–Piercing–Incendiary–Tracer M20,* U.S Army ARDEC, 1995.

[5] MESSINA,N.A., INGRAM,L.S., SUMMERFIELD,M.: *Design and fabrication of a prototype tracer surveillance tester*, US ARRADCOM, Defense Technical Information Center, ADA131612, 1983.

[6] BUC,S.M., ADELMAN,G., ADELMAN,S.: *Development of Alternate 7.62mm Tracer Formulations*, US ARRADCOM, Defense Technical Information Center, ADA335414, 1993.

[7] SHULMAN,R.S.: *Factors affecting small arms tracer burning*, Ftankford Arsenal, Defense Technical Information Center, ADA080279, 1956.

[8] INGRAM,B.V.: *Color Purity Measurements of Traditional Pyrotechnic Star Formulas*, Journal of Pyrotechnics, Issue 17 (2003) 1-18.

[9] SADEK,R., KASSEM,M., ABDO, M., ELBASUNEY, S: *Spectrally adapted red flare tracers with superior spectral performance*, Defence Technology, Volume 13, Issue 6 (2017) 406-412.

[10] SADEK,R., KASSEM,M., ABDO, M., ELBASUNEY, S: *Characteristics of the Red Colored Flame of Firework Compositions*, Journal of Pyrotechnics, Issue 28 (2009) 51-60.

[11] DAM,Q.S., NGUYEN,V.T., HOANG,T.H., NGUYEN,V.B.: *The preparation and characterization of the red flame composition based on strontium nitrate, magnesium, and polyvinyl chloride*, Journal of Science and Technique ,17(03) (2022) 62-70.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

RADIOACTIVE WASTE AS A LONG TERM SECURITY CHALLAENGE: PRINCIPLES OF PROTECTION AND RISK MANAGEMENT

MILICA ĆURČIĆ
Vinca Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, milica.curcic@vin.bg.ac.rs

PREDRAG BOŽOVIĆ
Vinca Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, bozovic@vin.bg.ac.rs

MARIJA JANKOVIĆ
Vinca Institute of Nuclear Sciences, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, marijam@vin.bg.ac.rs

Abstract: *This paper analyzes radioactive waste as a long-term security challenge through the prism of international protection principles and the national normative framework of the Republic of Serbia. Radioactive waste management encompasses a series of administrative and technical activities aimed at ensuring the long-term protection of people and the environment. Of particular importance are the standards of the International Atomic Energy Agency (IAEA), which introduce a graded approach based on risk assessment; therefore, safety and security requirements differ depending on the radioactive waste category and include administrative measures, physical protection, and institutional oversight, alongside the application of multiple engineered and natural barriers. The paper further examines the fundamental IAEA safety principles, such as responsibility, the role of the state, optimization of protection, and the protection of future generations, which constitute the basis for the sustainable management of radioactive waste. In normative terms, the legal framework of the Republic of Serbia is based on a specific legislative regime governing this area, supported by a number of bylaws and implementing regulations. However, challenges have been identified with regard to institutional coordination, strategic planning, and the integration of safety and security aspects. This paper points to the need for improving a comprehensive and long-term sustainable radioactive waste management system.*

Keywords: *radioactive waste, security, risk management.*

1. INTRODUCTION

Modern energy and technological developments have renewed focus on the question of nuclear energy's role in ensuring stable, low-carbon, and sustainable national energy systems. Alongside its potential benefits, the use of nuclear energy raises a range of complex safety, regulatory, and ethical issues, with radioactive waste management occupying a central place among them. Unlike many other types of industrial waste, radioactive waste poses a long-term potential risk to human health and the environment, necessitating a systemic approach based on long-term planning, institutional responsibility, and internationally harmonized protection standards.

In this context, IAEA has developed a set of fundamental safety principles [1] that form the basis of the global nuclear safety and security architecture. These principles encompass not only technical protection aspects but also establish a broader framework integrating responsibility for safety and security, the role of the state, risk management, the protection of present and future generations, and the need for transparency and public involvement in decision-making processes. Their application enables states to develop sustainable and reliable radioactive waste management systems that are resilient to technical, safety, security, and societal challenges.

A particular complexity is added to this issue by the fact that radioactive waste management extends beyond the timeframe of a single generation, thereby opening the way for considerations regarding intergenerational responsibility, long-term safety, and environmental protection [2]. In this regard, the safety and security approach to this problem entails not only preventing accidents and controlling radiation exposure but also establishing systems that remain reliable and functional over extended periods. This approach also involves addressing potential security threats, such as unauthorized access to radioactive materials or their misuse, thereby further underscoring the need to integrate safety and security aspects.

In the Republic of Serbia, the issue of radioactive waste management is gaining importance in light of recent strategic and regulatory changes, including the recognition of the importance of nuclear energy in the

Corresponding author: P. Božović;
E-mail: bozovic@vin.bg.ac.rs
DOI: 10.5937/OTEH26151C

Energy Development Strategy [3], the abolition of the Moratorium on the construction of nuclear power plants [4] with the adoption of amendments to the Law on Energy [5] and the preparation of a Preliminary Technical Study on the Peacetime Application of Nuclear Energy [6]. Although the regulatory framework in the field of radiation and nuclear safety and security is established, certain aspects of radioactive waste management remain fragmented, especially in relation to the general system of waste management and emergency situations. This situation indicates the need for further improvement of the normative and institutional framework, with the aim of harmonizing with international standards and ensuring a comprehensive approach to this issue.

Based on the above, the aim of this paper is to analyze the basic principles of radioactive waste management defined by the IAEA, as well as to consider their application and challenges in the context of the Republic of Serbia. Special emphasis is placed on the security aspects of this issue, including long-term risks, institutional responsibility and the need for an integrated approach that links regulatory, technical and social aspects of radioactive waste management.

2. INTERNATIONAL PRINCIPLES OF PROTECTION IN RADIOACTIVE WASTE MANAGEMENT

Radioactive waste management includes all administrative and operational activities related to the handling, pretreatment, treatment, conditioning, transport, storage and final disposal of radioactive waste, with the aim of ensuring long-term protection of people and the environment. In order to achieve this, it is necessary for radioactive waste to be adequately characterized, processed and disposed of in a scientifically based and ethically responsible manner [7].

According to IAEA standards, the safety of radioactive waste is implemented using a graded approach, where the level of physical protection (detection, delay and response) is based on potential hazards and the category of radioactive waste. The danger mentioned as potential is based on the D value of the radioactive material contained in the radioactive waste and is defined for each radioisotope individually. These values are shown in the IAEA Categorization of Radioactive Sources [8]. They represent a specific activity for which a given radioisotope can cause a severe deterministic effect if radioactive source is not under control. Based on this value and activity, sources are categorized from 1 to 5, where 1 is the highest category.

Based on the categorization of sources, according to the IAEA document [9], the safety requirements are defined through three levels A, B and C with the correlation of source category 1 - level A, source category 2 - level B and categories 3 to 5 - level C. On the other hand, the classification of radioactive waste basically serves to determine the required depth and type of disposal according to the time during which the material is considered radioactive. According to the IAEA document [10], the following classification of radioactive waste was adopted: Exempt waste (EW), Very Short-Lived Waste (VSLW), Very Low-Level Waste (VLLW), Low Level Waste (LLW), Intermediate Level Waste (ILW) and High-Level Waste (HLW). For waste categories up to VSLW, security measures include mainly administrative controls to prevent the unauthorized removal of material that is still considered to be radioactive. LLW and ILW require both physical protection (detection and delay system) to prevent malicious acts and unauthorized access during storage and institutional control measures based on active monitoring and regulatory oversight. The most stringent requirements in terms of security measures are required by HLWs that form an integrated safety and security system and a long-term storage strategy.

When both aspects are looked at in the end (classification of radioactive waste and D values of the radioactive materials that make up that waste) the safety requirements can be summarized as follows through Tables 1 and 2.

Table 1. Determination of safety levels based on the type of radioisotopes in radioactive waste and the classification of radioactive waste.

Used description quantity		D value	Classification of radioactive waste
Input parameters for analysis	Primary purpose	Safety and response in the event of an emergency	Disposal and long-term management
	Quantities used	Activity and D-value ratio	Activity concentration and half-life
	Categories	1 - 5 (1 – highest)	1 - 6 (HLW – highest)
Output parameters of analysis	Result/implication	Dictates requirements for physical protection	Dictates requirements for protective barriers and disposal depth

Ensuring the long-term stability and sustainability of the disposal system involves the application of multiple engineering and natural barriers, designed to prevent the release of radionuclides into the environment. These barriers work complementary, enabling passive security and reducing dependence on active surveillance over time. In addition, establishing institutional controls, maintaining comprehensive documentation, and implementing long-term management strategies are key elements in maintaining oversight and accountability over a long period of time.

Equally important are public participation and

transparency in decision-making processes, which contribute to building trust, strengthening social acceptability and respect for ethical aspects. All activities related to the disposal of radioactive waste must be synchronized with relevant national regulations and international legal instruments, thus ensuring a harmonized and responsible approach at the global level. Finally, the minimization of radioactive waste generation and the integration of waste management issues into all phases of the nuclear cycle - from production and use to final disposal - are the basic prerequisites for achieving long-term sustainability and efficiency.

Table 2. Technical security requirements depending on the security level

Radioactive source category	Security level	Primary goal	Security equipment requiremts - examples
1	A	Prevention of unauthorized removal	Detection: alarm system, CCTV, real-time monitoring, PTS Delay: high-security vaults, containers with heavy-duty protective barriers, multiple barriers Response: immediate tactical response to interrupt the act Security equipment requirements
2	B	Minimization of the probability of removal	Detection: alarm system Delay: access control (keys, ID cards, and biometrics) Response: rapid assessment and response that does not necessarily entail interrupting the act
3	C	Reduction of the probability of removal	Detection: daily checks and a simple sensor system Delay: basic physical barriers Response: notification of competent authorities upon confirming the loss of the source
4	C	Maintaining administrative control	Equipment: secure storage, periodic inventory checks, basic locking system to prevent access
5	C	Maintaining administrative control	Equipment: secure storage, periodic inventory checks, basic locking system to prevent access

In this regard, the fundamental safety principles developed by the IAEA constitute an internationally recognized basis for the safe management and disposal of radioactive waste. These principles [1], briefly outlined below, form the foundation for ensuring the long-term protection of human health and the environment, as well as the sustainability and transparency of radioactive waste management practices.

A. Responsibility for safety

Primary responsibility for safety rests with the license holder, that is, the natural or legal person authorized to conduct the activity or operate the facility involving radiation risks. This responsibility applies throughout the entire lifecycle of the facility or activity, regardless of the involvement of other actors. From a safety perspective, clearly defined and indivisible responsibility forms the basis for preventing institutional gaps and the shifting of responsibility during crisis situations.

B. Role of the State

The State is responsible for establishing and maintaining an effective legal and institutional framework for safety. This entails the formation of an independent regulatory body, as well as the establishment of mechanisms for oversight, emergency preparedness, and the control of all radiation sources and radioactive waste disposal. An effective state role is crucial for establishing a comprehensive system capable of addressing complex safety challenges and managing risks at the national level.

C. Leadership and management for safety

Organizations conducting activities involving radiation risks are required to establish effective leadership and an integrated management system that prioritizes safety, fosters a strong safety culture, and ensures continuous performance improvement and the application of lessons learned, in order to prevent safety from being compromised by competing demands.

D. Justification of facilities and activities

The justification of facilities and activities involving radiation exposure is based on the principle that the overall benefits must outweigh the potential radiation risks, taking into account all relevant health, economic, social, and environmental factors. Decisions regarding justification are often made at the highest level of government, particularly in the context of strategic decisions such as the introduction of a nuclear program, whereas in individual cases, the regulatory body plays a key role in assessing the acceptability of specific activities.

E. Optimization of protection

Protection measures must be optimized to achieve the highest reasonably achievable level of safety throughout the entire lifecycle of a facility or activity. A graded approach is applied, encompassing normal, abnormal, and

accident conditions, while taking into account the interdependencies between various phases, groups of exposed persons, and the different stages of radioactive waste management.

F. Limitation of individual risk

Radiation risk management must incorporate clearly defined dose and risk limits to ensure that no individual is exposed to an unacceptable level of harm. These limits represent an upper permissible value and must be complemented by the continuous optimization of protection. Establishing clear risk limits is crucial not only for protecting individuals but also for maintaining public trust in the institutions of the system.

G. Protection of present and future generations

Protection must encompass both people and the environment, in the present and the future, while taking into account long-term and trans boundary risks. Radioactive waste management must be conducted in a manner that does not impose an undue burden on future generations, through the implementation of safe, feasible, and environmentally acceptable solutions. This principle introduces the concept of intergenerational safety as a key element of strategic planning in the field of nuclear policy.

H. Accident prevention

All practical measures must be taken to prevent and mitigate nuclear and radiological accidents, with particular attention paid to preventing the loss of control over a nuclear reactor, a chain reaction, or a radiation source. A key principle in this area is "defense-in-depth," achieved through a combination of multiple successive and independent levels of protection, technical, organizational, and operational, such that a failure or error at one level does not lead to harmful consequences. This principle encompasses high-quality design and engineering solutions, the use of reliable control and monitoring systems, a strong safety culture, rigorous operational procedures, and pre-defined accident management plans. This approach ensures that no single technical, human, or organizational failure can cause significant damage, while combinations of failures that could lead to serious consequences remain extremely rare; this significantly enhances system resilience and reduces the likelihood of accident escalation.

I. Emergency Preparedness and Response

Organizations and competent authorities must establish and maintain effective mechanisms for preparedness and response to nuclear and radiological emergencies at local, regional, national, and, where agreed upon, international levels. The aim of these preparations is to minimize radiation risks associated with foreseeable incidents and, in the event of an accident, to implement practical measures to protect human life, health, and the environment. Plans include criteria for implementing protective measures, capabilities for protecting and informing personnel and the public, and periodic exercises to verify the effectiveness and readiness of the relevant institutions. In emergency situations, subject to informed consent, emergency response personnel may receive doses exceeding standard occupational limits, but only up to a pre-defined level.

J. Protective measures for existing or unregulated risks

Protective measures aimed at reducing existing or unregulated radiation risks must be justified and optimized. Such risks may arise in situations outside the oversight of regulatory bodies, including natural radiation sources (e.g., radon exposure in dwellings and workplaces), legacies of past human activities that were not adequately regulated (such as residual radioactive materials from mining operations), or the consequences of uncontrolled radionuclide releases into the environment. Every protective measure entails economic, social, and environmental costs, and sometimes risks to the individuals implementing them. A measure is considered justified only if the benefits significantly outweigh the potential risks and other negative consequences, while optimization seeks to achieve the greatest possible benefit relative to the resources invested.

In addition to the basic principles defined in the aforementioned IAEA Fundamental Safety Principles, radioactive waste management is further elaborated through specific standards defining operational and technical requirements. Notable among these is *Predisposal Management of Radioactive Waste* [11], which covers all steps in radioactive waste management from generation to disposal, including pre-treatment, treatment, conditioning, storage, and transport. Such waste can be generated during the commissioning, operation, and decommissioning of nuclear facilities; the use of radionuclides in medicine, industry, agriculture, research, and education; the processing of materials containing naturally occurring radionuclides; and processes for the remediation of contaminated areas. At the European level, radioactive waste management is governed by Directive 2011/70/Euratom [12], which obliges Member States to establish national programs for the safe management of radioactive waste and spent nuclear fuel. The aim of this regulation is to ensure the long-term protection of people and the environment in accordance with international standards, while harmonizing practices among Member States and ensuring transparent planning for all stages of the radioactive material lifecycle.

3. NORMATIVE AND INSTITUTIONAL FRAMEWORK IN THE REPUBLIC OF SERBIA

The regulatory framework for radioactive waste management in the Republic of Serbia is based on a clearly defined legal structure comprising primary and secondary legislation. Although the general waste management system is governed by the Law on Waste Management [13], its provisions do not explicitly apply to radioactive waste; this distinguishes the area as a separate and specific category requiring a dedicated regulatory approach.

Radioactive waste management is governed by the Law on Radiation and Nuclear Safety and Security [14], which serves as the fundamental legal act in this field and

establishes key obligations and operational principles. Pursuant to this Law, the Rulebook on Radioactive Waste and Spent Nuclear Fuel Management ("Official Gazette of the RS", No. 127/2021) was adopted, providing detailed regulations on the operational and technical aspects of management [15]. This rulebook prescribes mandatory measures that include avoiding the generation of radioactive waste or minimizing it to the greatest extent possible, preventing criticality, preventing unauthorized access and handling, ensuring controlled storage and disposal, maintaining records, and establishing an emergency response plan. At the same time, it mandates that all management procedures must prevent environmental contamination, harmful impacts on flora and fauna, and adverse effects of ionizing radiation on employees and the general public. It is particularly significant that the rulebook recognizes the need for an integrated approach when radioactive waste also possesses the properties of other hazardous materials, thereby requiring compliance with regulations governing hazardous waste management as well.

Although the Regulation on Security Measures for Nuclear Facilities and Nuclear Materials [16], adopted pursuant to Article 61, Paragraph 3 of the Law on Protection against Ionizing Radiation and on Nuclear Safety ("Official Gazette of the RS," Nos. 36/09 and 93/12), defines technical, physical, and organizational security measures, its legal framework is limited in certain respects compared to the current Law on Radiation and Nuclear Safety and Security. The Regulation applies to the security of radioactive material storage, including technical and organizational aspects, yet it does not fully encompass the broader regulatory framework and strategic provisions of the current law. The current law, which incorporates insights from international standards and harmonizes national regulations with IAEA guidelines, expands the competencies of regulatory bodies, brings a wider range of activities under supervision, and emphasizes inter-institutional coordination. In this context, the aforementioned regulation remains relevant for operational and technical procedures but does not fully cover the law's new provisions regarding strategic oversight, long-term nuclear material management, and systemic security policies.

Similarly, the Decree on Establishing the Nuclear Safety and Security Program [17], adopted pursuant to Article 4, Paragraph 1 of the Law on Protection against Ionizing Radiation and Nuclear Safety ("Official Gazette of the RS", Nos. 36/09 and 93/12), governs long-term plans and objectives regarding nuclear activities, including the system of control and physical protection of nuclear facilities and nuclear materials, in accordance with the standards of international organizations and assumed international obligations. The Decree serves as an operational and strategic framework for implementing the nuclear safety and security program and constitutes an integral part of the regulatory system defined by law. However, compared to the current law, the Decree remains primarily focused on defining the Program and plans, while failing to cover the detailed provisions of the new law regarding oversight, the coordination of regulatory bodies, and the comprehensive management of nuclear materials throughout the entire lifecycle of facilities. This highlights the need for the continuous adjustment of subordinate legislation to ensure full alignment with modern principles of nuclear safety and security and international standards.

Despite the existence of a regulatory framework, certain structural and functional gaps are evident that may pose a challenge to the development of a comprehensive radioactive waste management system in the Republic of Serbia. Above all, separating radioactive waste from the general waste management system, while justified given its specific nature, can lead to fragmentation of the regulatory framework and insufficient integration with broader environmental protection policies.

Coordination among various competent authorities poses a significant challenge, particularly in areas where competencies overlap, such as environmental protection, hazardous materials management, and emergency response. The Law on Disaster Risk Reduction and Emergency Management [18] regulates disaster risk reduction, prevention, the strengthening of resilience and preparedness among individuals and communities, the protection and rescue of people and material and cultural assets, the rights and obligations of citizens, legal entities, and state and local government bodies, as well as emergency management, civil protection operations, early warning systems, and international cooperation. The Law also provides for inspection oversight and regulates other matters relevant to the effective functioning of the disaster risk reduction system. It is specifically emphasized that protection against accidents involving radioactive and nuclear materials is subject to special regulations.

Specifically, business entities and other legal persons that use hazardous substances in their operations, or may come into contact with them, are required to take all necessary measures to prevent accidents and limit their consequences for human life and health, the economy, the environment, and social stability. This includes drafting an accident protection plan and submitting it to the Ministry for approval, as well as implementing the measures defined therein. Particular attention is paid to facilities for the exploitation of mineral resources and those where radioactive and nuclear materials are used, stored, or disposed of, as the consequences of technical or technological accidents at such sites can severely impact human life and health, economic and environmental systems, and social stability. In such cases, legal persons are obliged to organize and carry out protection and rescue operations for people, property, and the environment in accordance with specific regulations. Consequently, it is essential to further improve cooperation and information-sharing mechanisms among institutions, given that various laws stipulate different obligations and involve different institutions responsible for taking action. Such a coordinated approach would enable more effective implementation of preventive measures, better preparedness for potential incidents, and an overall strengthening of the system's safety and resilience.

Furthermore, although the law and accompanying bylaws regulate the technical and operational aspects of management in detail, the strategic dimension—encompassing long-term planning, the selection of permanent solutions for radioactive waste disposal, and a clearly defined national policy in this area—is less pronounced. This is particularly significant given current deliberations regarding the potential development of a nuclear program in the Republic of Serbia.

From a security and safety perspective, while regulations acknowledge the need for protection against unauthorized access and the definition of emergency measures, there is still room to further develop an integrated approach that fully links safety and security components, especially regarding the protection of radioactive materials against misuse. These gaps highlight the need to further develop an integrated, strategically oriented, and cross-sector aligned radioactive waste management system that is fully adapted to the potential future requirements of a nuclear program in the Republic of Serbia.

4. CONCLUSION

Radioactive waste management represents one of the most complex issues in modern safety and environmental policy, given its long-term nature, potential risks, and the need for sustainable solutions that extend beyond the timeframe of a single generation. Analysis shows that international standards, primarily those developed by the IAEA, provide a clear and comprehensive framework based on the principles of a graded approach, optimization of protection, responsibility, and the protection of future generations. These principles enable the establishment of a system that integrates technical, organizational, and institutional measures to minimize risk and ensure long-term safety and security.

The Republic of Serbia's regulatory framework in this area is built on a solid foundation, comprising a dedicated law and accompanying bylaws that recognize radioactive waste as a distinct category requiring a specialized regulatory approach. However, certain weaknesses have been identified, including system fragmentation, insufficient coordination among competent institutions, and an underdeveloped strategic management dimension, particularly regarding long-term disposal and planning in the context of a potential nuclear program. There is also room for further improvement in the integration of safety and security aspects.

Accordingly, the further development of the radioactive waste management system in the Republic of Serbia should focus on strengthening inter-institutional cooperation, enhancing the regulatory framework in line with the latest international standards, and developing long-term strategies that encompass the entire lifecycle of radioactive material. Enhancing transparency and involving the public in decision-making processes is of particular importance, as this contributes to building trust and social acceptance. Only an integrated, strategically oriented, and sustainable approach can ensure the effective management of radioactive waste while safeguarding human health, protecting the environment, and maintaining the long-term stability of the system.

Acknowledgement

The research was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Grant No. 451-03-33/2026-03/200017).

References

[1] International Atomic Energy Agency (2006). Fundamental Safety Principles, *IAEA Safety Standards Series No*. SF-1, Vienna.

[2] O. Korede, O. Akinyemi, Oshilalu, A. (2003). Environmental and ecological impact of radioactive waste disposal, *World Journal of Advanced Research and Reviews*, 18(02), 1419–1439.

[3] Energy Development Strategy of the Republic of Serbia until 2040 with projections until 2050 ("Official Gazette of the Republic of Serbia", No. 94 of November 28, 2024)

[4] Law on the Prohibition of the Construction of Nuclear Power Plants in the Federal Republic of Yugoslavia ("Official Gazette of the FRY", No. 12/95 and "Official Gazette of the RS", No. 85/2005 – other law).

[5] Energy Act ("Official Gazette of the RS", Nos. 145/2014, 95/2018 – other law, 40/2021, 35/2023 – other law, 62/2023, 94/2024 and 109/2025 – other laws).

[6] Ministry of Mining and Energy of the Republic of Serbia. (March 14, 2026). Preliminary technical study on the peaceful use of nuclear energy presented. https://www.mre.gov.rs/vest/sr/14964/predstavljena-preliminarna-tehnicka-studija-o-mirnodopskoj-primeni-nuklearne-energije.php

[7] International Atomic Energy Agency (2003). *Radioactive Waste Management Glossary, 2003 Edition*, IAEA, Vienna.

[8] International Atomic Energy Agency (2005). *Categorization of Radioactive Sources*, IAEA Safety Standards Series No. RS-G-1.9, Vienna

[9] International Atomic Energy Agency (2019). *Security of Radioactive Material in Use and Storage and of Associated Facilities*, IAEA Nuclear Security Series No. 11-G (Rev. 1), Vienna.

[10] International Atomic Energy Agency (2009). *Classification of Radioactive Waste*, IAEA Safety Standard Series No. GSG-1, Vienna.

[11] International Atomic Energy Agency (2009). *Predisposal Management of Radioactive Waste*. Vienna: IAEA (General Safety Requirements No. GSR Part 5).

[12] European Union. (2011). Council Directive 2011/70/Euratom of 19 July 2011 establishing a community framework for the responsible and safe management of spent fuel and radioactive waste. Official Journal of the European Union, L 199/48.

[13] Law on Waste Management ("Official Gazette of RS", Nos. 36/2009, 88/2010, 14/2016, 95/2018 – other law, and 35/2023).

[14] Law on Radiation and Nuclear Safety and Security ("Official Gazette of RS", Nos. 95/2018 and 10/2019 – other law)

[15] Rulebook on the Management of Radioactive Waste and Spent Nuclear Fuel ("Official Gazette of RS", No. 127/2021)

[16] Decree on Security Measures for Nuclear Facilities and Nuclear Materials ("Official Gazette of the RS", Nos. 36/09 and 93/12, Belgrade, 2009/2012)

[17] Decree on Establishing the Nuclear Safety and Security Program ("Official Gazette of the RS", Nos. 36/09 and 93/12, Belgrade, 2009/2012)

[18] Law on Disaster Risk Reduction and Emergency Management ("Official Gazette of RS", Nos. 87/2018, 14/2019 – other law, 71/2019 and 43/2021)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EXPLOATATION OF ARMAMENT & MILITARY EQUIPMENT-INTERACTION MAN AND TECHNOLOGY: ORGANIZATIONAL CULTURE AS THE FOUNDATION OF MILITARY STANDARDS IN THE MODERN MILITARY PROFESSION

LAZAR BJELICA
Strategic Research Institute, University of Defense, Belgrade, lazar.bjelica@mod.gov.rs

SRĐAN ŽIVKOVIĆ
Military Technical Institute, Belgrade, srdjan.zivkovic@mod.gov.rs

Abstract: *The accelerated progress and digitization of defense technologies imposes the need for a deeper understanding of the interaction of man and technology within the military profession. The paper sets up a conceptual framework for analyzing how organizational culture acts as a guarantor of consistent application of military standards and building trust in technical assets. Organizational culture is viewed as a set of internalized values and behavioral patterns that shape the attitude of system members towards technology, risk and prescribed procedures. Relying on the sociotechnical theory, the concept of highly reliable organizations and the theory of trust, the paper distinguishes key dimensions: perceived technical reliability, safety culture, the relationship between formal procedures and informal practices, as well as the impact of automation on the sense of control and decision-making. Special attention is paid to the role of technical systems and standards as active factors that shape work practice. In this way, it is indicated that the exploitation of military assets is not exclusively a technical, but also a sociocultural issue. The work is conceived as a theoretical analysis, with a proposal of a model of operationalization through examples of questions and questionnaires that have an illustrative function and indicate possible directions for future empirical research.*

Keywords: *exploitation, armament & military equipment, standards (technical & military), organizational culture.*

1. INTRODUCTION

The rapid advancement of modern technologies is witnessed on a daily basis. In recent decades, technological development in the defense sector has been characterized by accelerated digitalization, the introduction of complex sensor, information, and automated systems, as well as the networking of weapon platforms into unified systems. Recent and ongoing conflicts in Nagorno-Karabakh, Ukraine, and the Middle East have demonstrated that without rapid digitalization, the implementation of artificial intelligence (AI), and automated systems, effective deterrence of aggressors and the conduction of conflicts are no longer viable or economically sustainable.

The Serbian Armed Forces and the Ministry of Defence are rapidly embarking on the "Digital Army" project [1], which will significantly strengthen the defense posture of the Republic of Serbia while further propelling the entire system, organizational units, and enterprises operating within the defense sector. These structures, which encompass specialized military-technical institutions and defense industry factories, form the core of the Defence Technological and Industrial Base of Serbia (OTIBS), whose capacities in research, development, testing, quality control, and metrology hold strategic importance for national security and the system's response to contemporary threats [10]. The integration of advanced technologies, such as artificial intelligence, virtual, and augmented reality, into the military system does not merely constitute a technical and engineering modernization, but represents a reality, a necessity, and an imperative for the efficient functioning and survival of the organization in a contemporary information society [20]. This trend dictates the need to evaluate the functioning of defense systems not solely through engineering and technical parameters, but also through the manner in which members of the armed forces-soldiers, officers, engineers, and workers-interact with this technology on a daily basis. The exploitation of weapons and military equipment, understood in its broadest sense as the use, handling, and maintenance of technical assets in accordance with prescribed procedures, represents the convergence point of two vital aspects. One is technical, grounded in design and engineering solutions, while the other is social, built upon the norms, experience, skills, and behavioral patterns of the organization's members. In the sociological theory of organizations, it has long been recognized that the influence of culture on organizational behavior operates at a deep, subconscious level, generating thought patterns that are treated within the system as entirely natural.

When advanced technologies or external standards are introduced into military technical systems, it must be taken into account that the subconscious assumptions and values of the actors themselves exert a decisive influence

Corresponding author: L. Bjelica;
E-mail: lazar.bjelica@mod.gov.rs
DOI: 10.5937/OTEH26152B

on the formulation and selection of practical solutions to operational problems. Neglecting this aspect poses a risk of parochialism, wherein the universal applicability of technical concepts is assumed without sufficient awareness of the specific local and organizational context in which the technology will be exploited [22].

The baseline assumption of this paper is that the consistent application of military and technical standards, as well as the trust users place in technical assets, are not solely a product of the quality of the technology itself, but significantly depend on the organizational culture in which that technology is implemented. Organizational culture is understood as a system of assumptions, beliefs, values, and behavioral norms developed and adopted by the members of an organization through shared experience, manifested through symbols, and guiding their thoughts and actions [2]. It is a relatively stable and clearly defined set of core values, beliefs, shared knowledge, norms, and customs that shapes organizational behavior, cognitive frameworks, and directs all activities of individuals and groups within the organization.

The objective of this paper is to offer a conceptual framework for analyzing the relationship between organizational culture, trust in technology, and the application of standards within a military environment. To this end, the study draws upon three theoretical sources: primarily socio-technical theory, as well as the concept of high-reliability organizations and trust theory. Based on these sources, four analytical dimensions are identified: (i) perceived technical reliability, (ii) safety culture, (iii) the relationship between formal procedures and informal practices, and (iv) the impact of automation on the sense of control and decision-making. Their interconnectedness allows the exploitation of weaponry to be viewed as a socio-technical phenomenon rather than the mere execution of technically defined procedures.

This paper is theoretical and analytical in nature. In the final section, a framework for the operationalization of the presented dimensions is developed, complete with illustrative questions and a draft questionnaire. Its purpose is to indicate potential directions for future empirical research within the defense system and enterprises engaged in the production and exploitation of weapons and hardware, without claiming to represent a definitive or validated measurement instrument.

2. ORGANIZATIONAL CULTURE IN A MILITARY ENVIRONMENT

As previously noted, organizational culture is most frequently defined in the literature as a system of shared assumptions, values, and norms that organization members share and use to interpret their own experience, which are transmitted to new members as the correct way to perceive, think, and feel in relation to the problems the organization faces. Although definitions of this concept vary across authors, they share a common distinction between visible cultural elements—symbols, rituals, language, formal regulations—and deeper, often unspoken assumptions from which these visible elements emerge [4]. The author refines this division through three clearly defined levels: artifacts (visible organizational structures and processes), espoused values (official strategies, goals, and norms), and basic underlying assumptions (unconscious, deeply ingrained beliefs that actually guide behavior) [4]. In military-technical systems, formal standards and operational procedures primarily emerge as artifacts and espoused values. However, to truly guarantee safety, they must be cognitively transformed through practice and collective experience into the basic assumptions of the users themselves. In doing so, the standard ceases to be merely an external obligation and becomes an integral part of professional identity.

Organizational culture is a dynamic process that continuously evolves, is nurtured, and is internalized by transmitting established norms and standards of behavior from more experienced to junior colleagues. It is grounded in community and symbols through which employees share and interweave their values within the collective. Due to its transmissible nature, this culture evolves into a stable tradition of the enterprise or organization, while simultaneously maintaining the flexibility required to adapt quickly and easily to current trends and environmental needs.

In the context of contemporary organizational transformations, particularly driven by rapid technological development, the traditional bureaucratic model characteristic of military systems is undergoing structural changes. Classical bureaucratic structures that strictly rely on formal procedures and hierarchy exhibit severe limitations in dynamic environments. Consequently, there is a growing need to develop more flexible, decentralized forms that influence internal culture and the adoption of technological innovations [23].

Such a cultural framework exerts a decisive influence on the daily behavior of individuals, their productivity, as well as general job and organizational satisfaction. When values are clearly established, they serve as a powerful internal driver and motivator for employees, directing their daily efforts, engagement, and energy toward achieving the shared strategic goals and priorities of the organization [5]. In the context of complex military-industrial systems, the success of implementing a broader development vision directly correlates with the degree to which employees understand the organizational strategy; without this understanding, individuals cannot adequately link their daily operations and equipment handling to strategic outcomes [10]. The military organization represents a highly specific environment for the development and operation of organizational culture. A pronounced hierarchy, emphasized discipline, formalized procedures, and the existence of thoroughly prescribed standards, technical manuals, and operational regulations might create the impression that the behavior of personnel is entirely dictated by formal rules. However, formal procedures never fully encompass the complexity of real-world work scenarios. The space "between" regulations and reality is filled by informal behavioral patterns,

established habits, tacit rules, the ways experience is passed from senior to junior personnel, and the minor shortcuts developed during daily work. It is precisely these patterns that largely determine whether a prescribed standard will be consistently applied in a concrete situation.

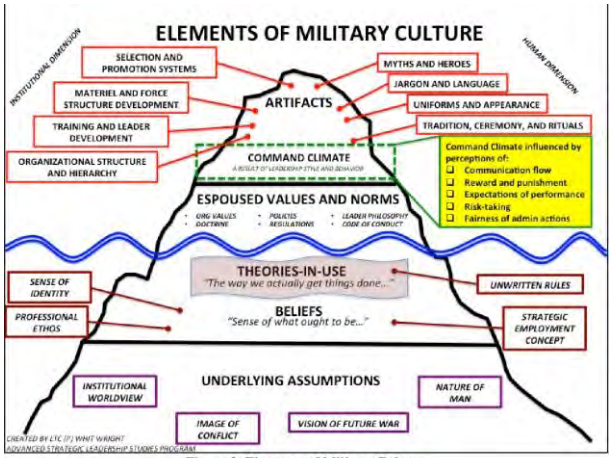


Figure 3. Elements of Military Culture

Figure 1. Elements of Military Culture [9].

For this reason, organizational culture is not treated in this paper as a secondary or less significant variable in relation to technology; rather, it is regarded as a factor that actively mediates between prescribed standards and the actual behavior of all users of technical equipment. In this way, culture becomes a mechanism that can either strengthen or undermine system reliability to an equal extent, regardless of the objective quality of the technology itself.

3. THEORETICAL FOUNDATIONS

The sociotechnical systems theory of organization, the foundations of which were established in the mid-twentieth century through studies of miners in British coal mines [6], posits that an organization's technical and social subsystems must be optimized jointly rather than in isolation. The introduction of a new technological solution inevitably alters the division of labor, communication flows, and power dynamics within work groups; consequently, a technically "optimal" solution may yield poor results in practice if it disrupts established social patterns of cooperation and trust. When applied to a military context, this perspective implies that the deployment of new weaponry is not solely a technical matter of procurement and training. Instead, it requires alignment with the existing unit structure, the division of responsibilities, and the established operational routines of the crew or team operating the system.

The concept of high reliability organizations (HROs) was developed through the study of systems operating under conditions of extreme complexity and potentially catastrophic consequences-such as nuclear power plants and operations on military aircraft carriers-which, despite these hazards, record a surprisingly low number of serious incidents [7]. This finding also served as a response to earlier, more pessimistic perspectives regarding the inevitability of severe accidents in complex, tightly coupled technical systems [8]. Researchers have identified five interrelated characteristics inherent to high reliability organizations:

- a persistent preoccupation with failure and a refusal to dismiss minor signs of disruption as insignificant;
- a reluctance to oversimplify complex events;
- a pronounced sensitivity to operations and the actual state of affairs "on the ground";
- a commitment to resilience, ensuring the system can recover rapidly following an error; and
- a deference to expertise, valuing the input of those closest to the source of risk regardless of their formal rank or stature [11].

Military units that handle complex and potentially hazardous weaponry belong, by the very nature of their mission, to this type of organization; thus, the question of how these characteristics are developed, nurtured, or, conversely, eroded through daily practice emerges as a legitimate research question.

In organizational theory, trust is most frequently examined through interpersonal relationships, where the assessment of the other party's ability, benevolence, and integrity stands out as key factors [12]. When it comes to trust in technology, one must also account for the fact that a technical asset possesses neither intent nor a moral component, yet the user nevertheless develops a specific type of "trust" in the system. Trust is of particular importance in situations where the complexity of an automated system is such that the user is unable to fully comprehend all of its processes and algorithms. Under such circumstances, trust serves as a mechanism that enables decision-making and reliance on the technical system under conditions of uncertainty [13]. It is the conviction that the system will behave in a predictable and reliable manner within the boundaries of its designed capabilities. Research in the field of automation indicates that the user's relationship with technology oscillates between two unfavorable extremes: over trust, where control and verification procedures are neglected because the system is deemed infallible, and under trust, where reliable technology is avoided or bypassed due to user skepticism [13]. Building "calibrated" trust, aligned with the actual capabilities and limitations of the system [13], represents a particularly sensitive issue in the military context, where both over trust and under trust in weaponry and equipment can have severe consequences for safety and successful mission execution.

Trust in automated systems within complex organizations is recognized as a multidimensional construct that, in addition to perceptions of reliability and competence, encompasses affective and experiential components. Recent empirical research on human-autonomy interaction demonstrates that user trust is not based solely on successful task execution or technical accuracy, but primarily on the quality of the interaction process and the manner in which the system manages uncertainty [19].

Safety culture encompasses shared beliefs and practices that determine how seriously risk will be treated in daily operations, independent of the moments when formal

control or inspection takes place. Key elements of a safety culture include a willingness to report minor deviations and occurrences without fear of punishment, the practice of learning from incidents rather than merely assigning accountability, and consistency between organizational rhetoric on safety and actual daily practices [14]. This logic closely mirrors the distinction between an approach that equates safety with the absence of negative outcomes and an approach that conceptualizes safety as an active capacity of the system to function correctly under most conditions, thereby requiring learning from examples of successful operations rather than exclusively from failures [15].

High-Reliability Organizations maintain their performance precisely through this type of collective safety and systematic learning [16]. In a military environment, where hierarchy and discipline are intricately intertwined with safety issues, it is particularly vital that the error-reporting culture does not conflict with the demands of discipline and adherence to the chain of command—specifically, a clear distinction must be made between negligent non-compliance with procedures and the honest admission of an oversight aimed at preventing future incidents.

When the previously outlined dimensions are viewed collectively, it becomes clear that the consistent and safe operation of weaponry and military equipment depends not solely on the technical serviceability of the asset and the formal completeness of regulations, but to a significant degree on how the organizational culture aligns the perception of reliability, safety practices, attitudes toward procedures, and responses to automation. The exact same technology, embedded within different organizational cultures, may in one instance be utilized precisely in accordance with the designer's intent and designated risk parameters, while in another, it may lead to severe deviations without a single formal modification to the weapon system itself. This supports the assertion that enhancing reliability and safety in the operation of military assets cannot be viewed exclusively through the prism of acquiring more advanced technology or enacting increasingly detailed regulations; rather, it demands parallel attention devoted to developing an organizational culture that supports the rigorous application of standards, appropriately calibrated trust in technology, and a sustainable balance between automation and human control.

Based on the presented theoretical framework, it can be hypothesized that organizational culture serves as a mediating factor between military standards, trust in technical systems, and operational safety. It is posited that organizations possessing a well-developed safety culture, consistent procedural compliance, and trust in technical systems achieve a higher degree of operational reliability and a lower risk of errors driven by the human factor. By integrating the specificities of the domestic defense environment into this theoretical framework, it becomes evident that strategic leadership must define an appropriate model for organizational culture change within the Serbian Armed Forces, aiming to create the necessary conditions for the development of operational capabilities and the efficient employment of forces in a dynamic and rapidly changing environment [21].

4. FUTURE EMPIRICAL RESEARCH

Although the paper is conceived as a theoretical-analytical study, the presented conceptual model opens up avenues for future empirical research within the defense system. Grounded in the identified theoretical constructs, it is possible to develop a standardized survey questionnaire to investigate the relationship between organizational culture, the application of military and technical standards, trust in technical systems, and the operational safety of weaponry and military equipment.

The prospective research is quantitative in nature and relies on an anonymous survey of defense system personnel directly involved in the design, testing, metrology, maintenance, operation, and management of technical assets. Respondents comprise military and/or civilian personnel, including engineers, researchers, and technical staff participating in these processes within military institutions (such as VTI, TOC, TRZ...) or defense industry factories.

The introductory section of the questionnaire covers the sociodemographic and professional characteristics of the respondents, such as gender, age, defense system status, field of work, and years of service. These data enable the analysis of potential attitudinal variations among different professional groups and levels of work experience. This approach to sample segmentation is scientifically grounded in existing studies of military systems, which demonstrate statistically significant differences in safety perception and readiness to modify practices across different service branches (e.g., more pronounced safety improvement practices in the air force compared to the army). This variance is attributed to the specific operational environments where errors in technical utilization carry immediate, catastrophic consequences [17].

The core section of the questionnaire is organized into several thematic blocks designed in alignment with the theoretical framework developed in this paper. Each of these blocks comprises items directly linked to key concepts and research objectives, ensuring the systematic collection of relevant data and facilitating subsequent analysis in accordance with the established theoretical premises.

The first domain addresses attitudes toward military and technical standards, metrology, and maintenance procedures. The objective is to examine the extent to which personnel accept standards as an essential element of the safety and reliability of technical systems. One of the items reads: "*During the operation or maintenance of an asset, I consistently rely on official technical manuals and prescribed procedures*."

The second thematic area explores trust in technology, digitalization, and automation. This segment focuses on assessing trust in modern technical systems, the adoption

of digital technologies, and the impact of automation on decision-making. An illustrative item states: *"The introduction of new technologies contributes to increasing occupational safety."*

The third domain encompasses organizational culture and safety culture, specifically focusing on procedural compliance, the role of leadership, the treatment of errors, and the transfer of professional knowledge. For instance: "*Operational errors are analyzed systematically, with the aim of learning rather than individual punishment*." These and similar questionnaire items allow for an assessment of the degree to which the organization exhibits the characteristics of High-Reliability Organizations. This segment builds directly upon modern methodological modifications of military surveys, according to which reflection on the impact of regulations on operations and the utilization of incident reporting are treated as primary tools for improving safety procedures at the tactical unit level [17]. Developing these items tests the openness of the military environment to feedback loops, which are critical for adaptive command and control in complex sociotechnical systems [17]. Within this domain, it is justified to introduce specific items measuring the safety leadership capacity of superiors. Validation for this addition lies in the fact that, within military structures, command personnel (officers and non-commissioned officers) typically manifest a significantly higher level of safety knowledge and safety communication compared to enlisted soldiers. Leadership in a military context thus represents the primary channel through which formal procedures and technical operation standards are translated into the conscious behavior of subordinates [3].

The fourth thematic area relates to the perception of technical reliability and risk. It examines how employees assess the risk of working with technical equipment, the extent to which they trust the outcomes of technical inspections, and how they react when results deviate from expectations. One such item is: "*Trust in prescribed procedures reduces my fear of asset malfunction or failure*." Evaluating this dimension is of particular importance for reference military-technical centers engaged in weapons testing, where maintaining objective independence in decision-making regarding test results constitutes a critical prerequisite for building end-user trust in the reliability of military assets [10].

The fifth domain concerns ethics and professional responsibility. The intention is to determine the degree to which metrology, technical serviceability, and compliance with procedures are internalized as a part of professional responsibility, rather than viewed merely as a formal obligation. An example item is: "*How would your confidence in crisis situations be affected by the knowledge that the asset had not undergone regular metrological inspection*?"

In addition to closed-ended questions utilizing a five-point Likert scale, the questionnaire would contain a small number of multiple-choice categorical questions, as well as several open-ended questions allowing respondents to share their own observations regarding the factors that influence trust in technical assets, the greatest operational challenges, and opportunities for enhancing the maintenance and quality control systems.

In this manner, the assumptions set forth in this paper are empirically tested, and the interrelationships among organizational culture, the application of standards, trust in technical systems, and the operational safety of weaponry and military equipment are analyzed. The obtained results serve as a foundation for improving the training system, developing organizational culture, revising maintenance procedures, and further advancing the digitization process within the defense system.

5. CONCLUSION

This paper proceeds from the premise that the operation of weaponry and military equipment constitutes a complex sociotechnical process in which the technical characteristics of assets, human behavior, and organizational conditions are mutually interdependent. Theoretical analysis indicates that the consistent application of military and technical standards, as well as the cultivation of adequate trust in technical systems, do not depend solely on the quality of technology or formally prescribed procedures, but to a significant degree on the organizational culture within which these assets are utilized. Drawing upon sociotechnical theory, the framework of High-Reliability Organizations, trust theory, and the concept of safety culture, this study proposes a conceptual framework that facilitates an integrated examination of the relationship between humans, technology, and organization within the military environment. Crucially, it is highlighted that organizational culture acts as a mediating factor influencing the consistency of procedural compliance, the adoption of new technologies, risk response mechanisms, and the development of trust in technical systems.

The proposed model for future empirical research does not constitute a definitive measurement instrument; rather, it demonstrates how the outlined theoretical concepts can be operationalized and empirically validated. Research based on this model would enable the determination of the interrelationships among organizational culture, the application of standards, trust in technology, and operational safety, while also identifying areas where organizational and technical improvements can be achieved. The scientific contribution of this paper lies in bridging concepts that are traditionally examined in isolation within the literature, and in demonstrating that the operation of weaponry and military equipment cannot be viewed exclusively as an engineering process. Instead, it must be understood as the product of interaction between technical systems, organizational values, and the professional behavior of their users.

Contemporary trends and security challenges necessitate a transition toward more flexible and operationally integrated workflow management systems. This implies a reliance on smaller organizational units comprised of highly qualified individuals, a reduction in management

layers, and the realization of an authentic, organic convergence between the technical and social subsystems within the defense system [18]. Furthermore, synthesizing social aspects and applying them to technical disciplines would yield a significant contribution, fostering enhanced collaboration between the social and technical sciences, particularly within the military context. Concurrently, this paper serves as a baseline for future empirical investigations that could contribute to optimizing training systems, cultivating organizational culture, rigorously applying standards, and further advancing digitization processes within the defense system. Organizational culture integrates technical standards into the military ethos and professional identity. Thus, armament exploitation ceases to be a mere routine and becomes a direct expression of responsibility in the modern military profession.

Acknowledgements

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No 451-03-34/2026-03/200325).

References

[1] https://www.politika.rs/sr/clanak/732344/srbija-dobija-digitalnu-armiju , accessed: June 20, 2026.

[2] JANIĆIJEVIĆ, N. Organizational Behavior [*Organizaciono ponašanje*]. Datastatus, Belgrade, 2008, 301.

[3] HU, S. *Analysis of the Impact of Military Organization's Safety Culture on Safety Behavior: Focusing on the Mediating Effect of Safety Leadership*. Journal of Advances in Military Studies, 2020, 3(2), 63–8.

[4] SCHEIN, E. H.; *Organizational Culture and Leadership* (4th ed.). San Francisco: Jossey-Bass. 2010.

[5] LOJIĆ, R. *Organizational Culture [Organizaciona kultura*]. Vojno delo, 2010, 260–261.

[6] TRIST, E. L., & BAMFORTH, K. W. *Some Social and Psychological Consequences of the Longwall Method of Coal-Getting*. Human Relations, 1951, 4(1), 3–38.

[7] LA PORTE, T. R., & CONSOLINI, P. M. *Working in Practice but Not in Theory: Theoretical Challenges of High-Reliability Organizations*. Journal of Public Administration Research and Theory,1(1), 1991, 19–47.

[8] PERROW, C. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books. 1984.

[9] WRIGHT, J.W. *The Challenges of Adopting a Culture of Mission Command in the US Army*. School of Advanced Military Studies, Fort Leavenworth, Kansas, 2015., 17.

[10] PETROVIĆ, V., ZRNIĆ, B., ŽIVKOVIĆ, S. One vision of defense industry performance management of the Republic of Serbia. In: *Proceedings of the 8th International Scientific Conference on Defensive Technologies (OTEH 2018)*, Vojnotehnički institut, Belgrade, 2018, 577-583.

[11] WEICK, K. E., SUTCLIFFE, K. M., & OBSTFELD, D. *Organizing for High Reliability: Processes of Collective Mindfulness*. Research in Organizational Behavior, 21, 1999, 81–123.

[12] MAYER, R. C., DAVIS, J. H., & SCHOORMAN, F. D. *An Integrative Model of Organizational Trust*. Academy of Management Review, 20(3), 1995, 709–734.

[13] LEE, J. D., & SEE, K. A. *Trust in Automation: Designing for Appropriate Reliance*. Human Factors, 46(1), 2004, 50–80.

[14] REASON, J. *Managing the Risks of Organizational Accidents*. Aldershot: Ashgate. 1997.

[15] HOLLNAGEL, E. *Safety-I and Safety-II: The Past and Future of Safety Management*. Farnham: Ashgate. 2014.

[16] WEICK, K. E., & SUTCLIFFE, K. M. *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. San Francisco: Jossey-Bass. 2001.

[17] SCHÜLER, M. *Expanding the Military Safety Climate Questionnaire (MSCQ)*. In: Proceedings of the 30th International Command and Control Research and Technology Symposium (ICCRTS 2025), Swedish Defence University, 2025.

[18] MILEKIĆ, M. *Sociotechnical systems [Socio-tehnički sistemi]. In: Proceedings of the 7th Scientific-Professional Conference with International Participation "QUALITY 2011" [Zbornik radova 7. Naučno-stručnog skupa sa međunarodnim učešćem "KVALITET 2011"],*Neum, B&H, 2011.

[19] HERON, S., & LAU, M. C. *Trust in Autonomous Human-Robot Collaboration: Effects of Responsive Interaction Policies* (An In-Person Pilot Study). 2026.

[20] VULETIĆ, D. *The Significance of New Technologies in the Military Education of Officers [Značaj novih tehnologija u vojnom obrazovanju oficira]*. National Security Policy [Politika nacionalne bezbednosti], Year XVI, Vol. 30, No. 3/2025.

[21] MILKOVSKI, V. *Organizational Culture in the Serbian Armed Forces [Organizaciona kultura u Vojsci Srbije*]. National Interest [Nacionalni interes], Year XVI, Vol. 39, No. 3, 2020.

[22] MOJIĆ, D. *Organizations and National Culture [Organizacije i nacionalna kultura*]. Sociologija, Vol. XLIX, No. 4, 2007.

[23] MOJIĆ, D. *Organizations and (Post)modern Society [Organizacije i (post)moderno društvo]*. Čigoja štampa and the Institute for Sociological Research of the Faculty of Philosophy in Belgrade, Belgrade, 2014, 34.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

THE INFLUENCE OF ULTRASONIC CAVITATION ON THE SURFACE ROUGHNESS OF LASER-SINTERED MS1 STEEL

PAVLE LJUBOJEVIĆ
Faculty of Mechanical Engineering, Belgrade, pljubojevic@mas.bg.ac.rs

ŽARKO MIŠKOVIĆ
Faculty of Mechanical Engineering, Belgrade, zmiskovic@mas.bg.ac.rs

TATJANA LAZOVIĆ
Faculty of Mechanical Engineering, Belgrade, tlazovic@mas.bg.ac.rs

Abstract: *Cavitation is a very common phenomenon in mechanical systems involving the interaction between a solid body and a fluid acting either as a working medium or as a lubricant. Under certain operating conditions, the occurrence of cavitation leads to cavitation erosion, i.e., material degradation. Numerous experimental studies have demonstrated that cavitation erosion affects the mechanical properties of material. However, an even more pronounced effect is observed in the alteration of surface roughness.This phenomenon is not necessarily detrimental, as it has been reported in the literature that cavitation can be effectively utilized for polishing geometrically complex components, which are often produced using additive manufacturing technologies.In this study, the influence of cavitation on the surface roughness of samples fabricated by laser sintering of MS1 steel was investigated. Cavitation was experimentally induced under laboratory conditions according to ASTM G32 standard.. The samples were also subjected to post-processing treatments, including heat treatment (aging) and mechanical strengthening (shot peening), enabling an analysis of the effects of different strengthening methods and their combinations on surface roughness evolution. The measurements were performed at multiple time intervals in order to monitor the progression of cavitation effects over time.*

Keywords: *cavitation, cavitation erosion, surface roughness, MS1, additive manufacturing*

1. INTRODUCTION

Cavitation is a phenomenon that occurs in fluids when the pressure drops below the saturation pressure or when the temperature rises above the evaporation temperature. In mechanical systems, cavitation is primarily caused by pressure variations, which may result from local drops due to changes in geometry and fluid flow velocity (hydrodynamic cavitation), or from pressure fluctuations induced by vibrations (acoustic cavitation). [1–4] These two mechanisms are often coupled, since changes in geometry and vibrations are common in mechanical systems, leading to hydrodynamic–acoustic cavitation (HAC). [5]

Cavitation bubbles implode, generating strong shock waves in the case of homogeneous nucleation, or microjets in the case of heterogeneous nucleation, which damage material surfaces. The implosion (condensation) process occurs within an extremely short time frame, while the number of bubbles is large, so this type of material damage can be considered a dynamic load from the perspective of load-bearing capacity studies. The resulting damage, known as cavitation erosion, affects the properties of the material's surface layers, as demonstrated by numerous studies conducted on different materials. [6, 7] Above all, cavitation has a significant impact on surface roughness, which is a crucial surface parameter and directly affects the functionality of machine parts.

This phenomenon is often neglected in vital machine parts, and there is very little research on the impact of cavitation erosion on their operational capacity. Furthermore, this is a very common occurrence in vital machine elements such as bearings, gears, seals, etc. In the calculation standards for these elements, cavitation erosion is listed as a mode of destruction, but it is deemed to have no effect on operational capacity. Although it is not a dominant phenomenon compared to primary damages, material particles suspended in oil can secondarily cause abrasive wear; in systems whose operating parameters are such that cavitation is pronounced, this can significantly affect operational capacity and reliability. [8]

Since hydrodynamic cavitation is more common, optimization of geometry is possible in order to reduce cavitation, but such optimization must be carried out in accordance with other potential damage mechanisms in mechanical components. Recently, mainly for techno-economic reasons, mechanical parts are increasingly being manufactured using additive technologies. These technologies enable the rapid production of components directly from 3D models and successfully overcome the challenges associated with manufacturing parts of

Corresponding author: P. Ljubojević;
E-mail: pljubojevic@mas.bg.ac.rs
DOI: 10.5937/OTEH26153L

complex geometry. For this reason, it is essential to determine the behavior of 3D-printed materials under cavitation erosion conditions. In this study, the influence of cavitation erosion on the surface roughness of MS1 steel samples produced by Direct Metal Laser Sintering (DMLS) technology was investigated, with and without subsequent thermal and mechanical surface treatments.

2. MATERIAL AND TECHNOLOGY

The test samples were produced from MS1 steel powder, which corresponds to tool steel grade with European classification 1.2709, the American 18% Ni Maraging 300, and the German X3NiCoMoTi 18-9-5. According to the powder manufacturer's specification (EOS GmbH), the chemical composition of the steel is presented in the following table. [9]

Table 1. Chemical composition of MS1 powder

Element	%	Element	%
Ni	17 - 19	Cu	0.5
Co	8.5 – 9.5	C	0.03
Mo	4.5 – 5.2	Mn	0.1
Ti	0.6 – 0.8	Si	0.1
Al	0.05 – 0.15	P	0.01
Cr	max. 0.5	S	0.01
Fe		Balance	

The particle size of the powder ranges from 15 to 65 µm and the density of the manufactured part 8 g/cm³ The test samples are cylindrical in shape, with a height of 5 mm and a diameter of 10 mm.

The test samples were manufactured using Direct Metal Laser Sintering (DMLS) technology, which belongs to the Powder Bed Fusion (PBF) process category. Fabrication of the test samples was carried out on an M280 DMLS system. The chamber is equipped with a build plate measuring 250 × 250 mm, allowing a maximum build height of 325 mm. The printing was performed in the vertical direction, directly on the plate without the use of support structures. Each layer was deposited at a thickness of 40 µm. The laser operates at a power of 200 W, emitting infrared radiation with a wavelength of 1064 nm. Prior to the start of the process, the build plate was preheated to 40 °C to ensure stable processing conditions. [10, 11]

One group of the test samples was tested in the as-built condition, without any subsequent thermal or mechanical treatments. Since the manufacturer recommends aging of this steel to improve mechanical properties, particularly strength, another set of samples was subjected to aging according to the manufacturer's guidelines. To achieve the maximum guaranteed hardness of 54 HRC, the aging process was carried out for 6 hours at 490 ±10 °C, followed by air cooling. One of the objectives of this study was to compare the influence of aging and shot peening on the cavitation resistance of the samples. Therefore, one group of test samples was shot peened. In addition, an analysis of surface roughness was performed on specimens that were first aged and subsequently shot peened.

3. EXPERIMENTAL PROCEDURE

The experiment was conducted in accordance with ASTM G32. At defined time intervals, the specimens were exposed to cavitation, subsequently dried, and evaluated for mass loss and surface roughness. Since the focus of this study is exclusively on changes in roughness, mass loss measurements are not discussed. The ultrasonic cavitation device used is shown in Figure 1. The test sample was immersed in a water bath, positioned 0.5 mm below the horn tip. The water temperature was 25ºC. The horn, excited by mechanical vibrations generated through a piezoelectric element converting electrical energy, oscillated with an amplitude of 5 µm at a frequency of 20 kHz. These vibrations induced pressure fluctuations in the water layer between the horn and the specimen, leading to bubble formation and cavitation erosion. [12]

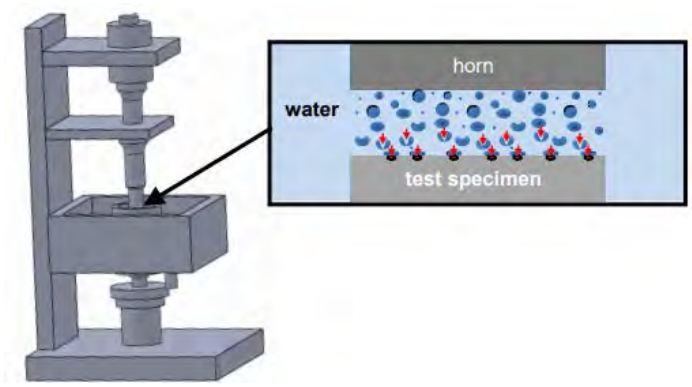


Figure 1. Device for ultrasonic cavitation testing

At specific measurement intervals, the surface roughness of the specimens was evaluated alongside mass measurements. For the as-built and aged specimens, roughness was assessed after 60, 120, and 240 minutes of cavitation exposure, whereas for the shot-peened and aged + shot-peened specimens, roughness was measured only after 240 minutes. Prior to cavitation testing, the initial roughness of each test sample was recorded. Roughness measurements were performed using a MarSurf SD26 device (Figure 2), which employs a diamond-tipped stylus. The stylus was positioned so that its tip lightly contacted the specimen surface. A cutoff length of 0.8 mm was applied, with a traversing length of 5.6 mm, and evaluation length of 4mm.

Figure 3 shows the surface roughness of the aged specimen after 240 min of cavitation erosion testing as an example of the measurement results. However, this study will not discuss the entire measured roughness profile; instead, the focus will be placed on the parameter Arithmetic Mean Height of the Assessed Profile, i.e., Arithmetic Mean Roughness – R_a. The change in the R_a parameter directly affects the operating capability of mechanical components exposed to cavitation, particularly gears, where surface roughness plays an important role in defining contact conditions and represents one of the parameters considered in the

correction factors for critical stress evaluation.

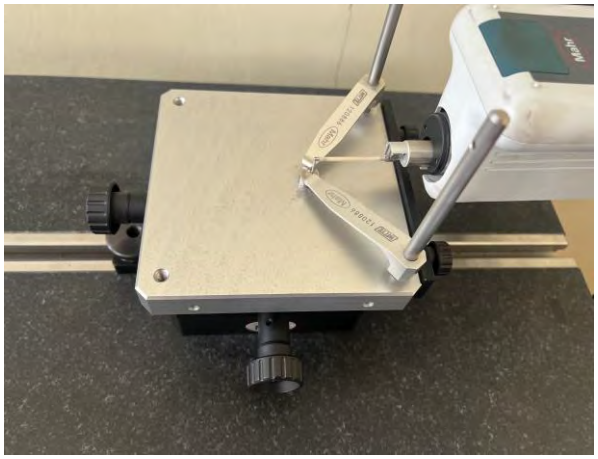


Figure 2. Measurement of test sample surface roughness

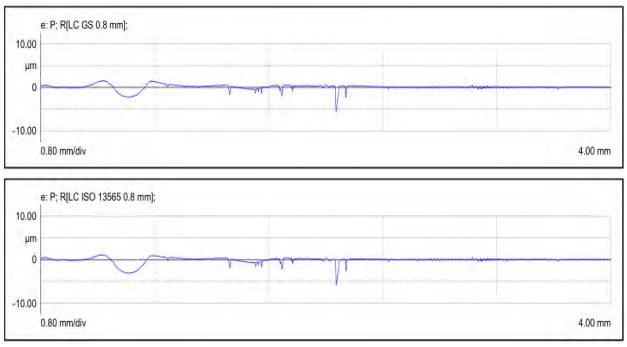


Figure 3. Surface roughness of aged MS1 steel after 240 min of cavitation exposure

4. RESULTS

Based on the measured surface roughness values for the as-built and aged samples, together with the initial roughness values, the dependence shown in Figure 4 is obtained.

One of the key parameters for the analysis of cavitation erosion is the Mean Depth of Erosion (*MDE*), which represents the average erosion depth over time and is calculated according to the following equation:

$$MDE_{\Sigma_i} = \sum_{i=0}^{n=8} \Delta MDE_i \sum_{i=0}^{n=8} \frac{4\Delta m_i}{\rho \pi d^2} 10^3 [\mu m] \quad (1)$$

Where: i is a testing period, ρ is a density, Δm_i is a mass loss measured in mg, and d is a test sample diameter.

The values of *MDE* were measured and obtained for the as built samples, as presented in the study [13]. Since both *MDE* and R_a represent averaged values, and experimental analyses have shown that the best approximation of *MDE* is achieved using the least squares method, in this work the values of R_a were also approximated using the least squares method. In the case of the aged samples, the mean squared error (MSE) is negligible, whereas for the as-built samples, it is comparatively higher.

Based on the obtained results for MDE, a straight line of the following form is derived with negligible error (R^2 = 0.9886):

$$MDE = 0.0105t - 0.0368 \quad (2)$$

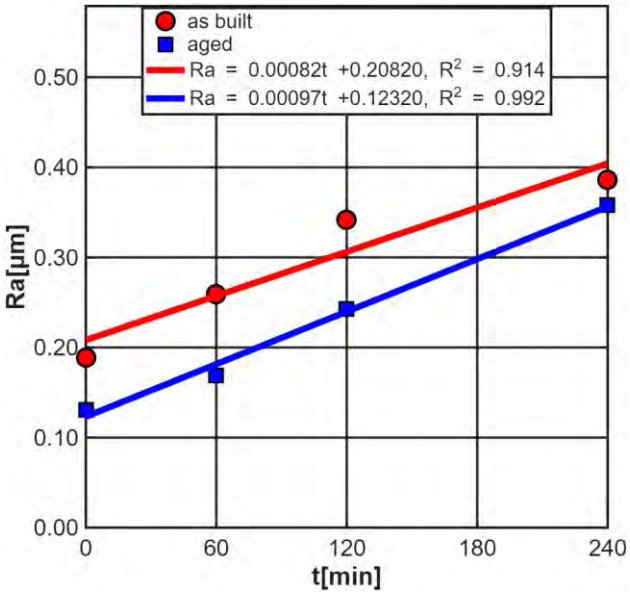


Figure 4. Surface roughness vs. time for as-built and aged MS1 steel

From the slope of the linear trends for R_a and *MDE*, it is evident that *MDE* changes at a considerably faster rate than R_a. A possible correlation between these two parameters could be established by comparing the lines without intercepts on the y-axis. In such a case, the variation of R_a could be predicted solely from its initial value and the time-dependent change of *MDE*. To define this relationship, a series of experiments would be required to determine the ratio between the slopes of the *MDE* and R_a curves. This ratio could then be expressed as a coefficient, enabling the prediction of one parameter based on the evolution of the other. Verification of this assumption would necessitate further experimental work and potentially the development of a mathematical model describing surface damage and roughness evolution. The obtained results also indicate that both *MDE* and R_a can be represented as linear functions of time. In that case, R_a could be calculated as follows:

$$Ra\,(t) = Ra\,(t = 0) + \frac{k_{MDE}}{C}\, t \quad (3)$$

Where C is the ratio of k for *MDE* and R_a. In this case of as-built MS1 steel, that ratio is 12.8.

For the test samples subjected to shot peening and combined aging + shot peening, surface roughness was measured at the beginning of the test and after 240 minutes of cavitation exposure. The results are presented in Figure 5. "Due to having only two measurement points, no linear regression approximation can be used, so the results are presented in the form of bars. The initial R_a values are significantly higher than in the case of these samples.

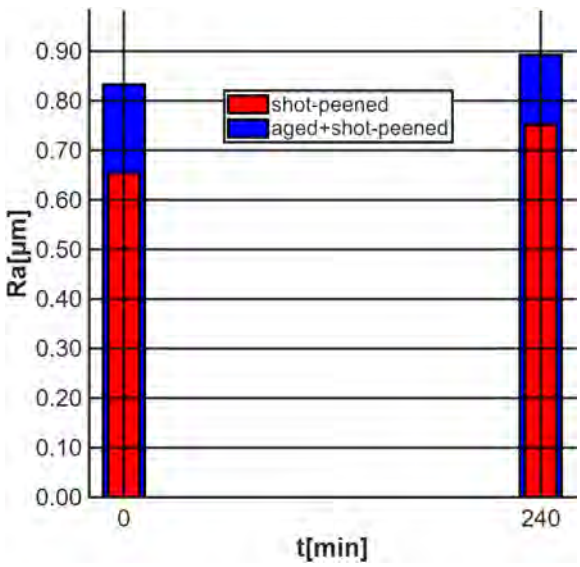


Figure 5. Surface roughness vs. time for shot peened and aged + shot peened MS1 steel

5. CONCLUSIONS

Based on the obtain results it can be concluded:

MS1 steel, characterized by excellent properties as a high-hardness steel, demonstrates very good performance in terms of surface roughness evolution. It exhibits only a minimal increase in R_a over time. Since R_a represents the average surface roughness, by analogy with MDE, which is the average depth of erosion, it can be approximated using the least squares method. The slope of this line is one of the indicators of the material's sensitivity to changes in surface roughness. These slopes are very small: 0.00082 for the as-built condition, 0.00097 for the aged condition, 0.00041 for the shot-peened condition, and 0.00025 for the aged + shot-peened condition (if a straight line were drawn through these two points), which indicates a very minor change in the R_a parameter over time.

For the MS1 as-built samples, analysis of the time-dependent changes in *MDE* and R_a indicates that *MDE* is a significantly more sensitive parameter than R_a, as its slope is nearly 13 times greater. By conducting a large number of experiments, it would be possible to establish a correlation between the time-dependent variations of *MDE* and R_a, enabling the prediction of one parameter based on the evolution of the other. For the as built and aged test samples, the slope of the line is higher than for the shot-peened and aged+shot-peened samples. One possible reason is the significantly lower initial roughness of the as built and aged samples, which were polished prior to testing and are very similar (as built – R_a = 0.189 µm, aged – R_a = 0.131 µm). An inconsistency observed is that the slope of the line approximating Ra for the aged samples is somewhat higher than for the as built samples, which should not be the case since aged samples exhibit greater cavitation resistance due to the higher surface hardness. However, the considerably larger approximation error in the as built samples likely influences this phenomenon. The initial roughness of the shot-peened and aged+shot-peened samples is significantly higher (shot-peened – R_a = 0.654 µm and aged+shot-peened – R_a = 0.833 µm), which is explained by the subsequent mechanical surface treatment (shot peening) that increases roughness. In addition to the strengthened surface, the initial roughness contributes to the gentler slope of the approximating curves in this case. Accordingly, as expected, the aged+shot-peened test samples exhibit the lowest slope of the line due to the highest surface resistance resulting from both thermal and mechanical treatment. In the case of the shot-peened and aged+shot-peened samples, only the initial and final roughness values were measured. Therefore, future experiments should include measurements at additional intermediate points to achieve a more precise determination of the slope of the approximating line.

References

[1] DOJČINOVIĆ,M.: *Razaranje materijala pod dejstvom kavitacije*, Monograph, Tehnološko-metalurški fakultet, Beograd, 2013.

[2] FRANC,J-P., MICHEL,J-M.: *Fundamentals of Cavitation*, Springer Science & Business Media, 2006.

[3] LAUTERBORN,O., OHL,C-D.: *Cavitation bubble dynamics*, Ultrasonics Sonochemistry, (1997)..

[4] NEPPIRAS,E.A.: *Acoustic cavitation series: part one*, Ultrasonics, 22(1) (1984) 25–28.

[5] WU,P., BAI,L., LIN,W., WANG,X.: *Mechanism and dynamics of hydrodynamic-acoustic cavitation (HAC)*, Ultrasonics Sonochemistry, 49 (2018) 89–96.

[6] ZAKRZEWSKA,D.E., KRELLA,A.K.: *Cavitation Erosion Resistance Influence of Material Properties*, Advances in Materials Science, 19(4) (2019) 18–34.

[7] HATTORI,S., ISHIKURA,R.: *Revision of cavitation erosion database and analysis of stainless steel data*, Wear, 268(1) (2010) 109–116.

[8] LJUBOJEVIĆ,P., LAZOVIĆ,T., DOJČINOVIĆ,M.: *Cavitation in Machine Elements: A Critical Review of Cavitation Damage, Experimental Methods, Standardization Challenges, and Applied Digital Technologies*, Lubricants, 14(6) (2026).

[9] EOS: *MDS EOS MaragingSteel MS1*.

[10] CROCCOLO,D., DE AGOSTINIS,M., FINI,S., OLMI,G., ROBUSTO,F., ĆIRIĆ KOSTIĆ,S., VRANIĆ,A., BOGOJEVIĆ,N.: *Fatigue Response of As-Built DMLS Maraging Steel and Effects of Aging, Machining, and Peening Treatments*, Metals, 8(7) (2018) 505.

[11] WONG,K.V., HERNANDEZ,A.: *A Review of Additive Manufacturing*, International Scholarly Research Notices, 2012(1) (2012) 208760.

[12] G02 COMMITTEE: *Test Method for Cavitation Erosion Using Vibratory Apparatus*.

[13] LJUBOJEVIĆ,P., LAZOVIĆ,T., DOJČINOVIĆ,M.: *Cavitation Erosion Parameters of Laser Sintered MS1 Steel Tested According to ASTM G32 Standard,* Proceedings IRMES2025, Vrnjačka Banja, 2025.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

NATURAL AND ANTHROPOGENIC FACTORS INFLUENCING THE VELIKA MORAVA RIVER WATER REGIME

MILAN KRESOJEVIĆ
University of Defence in Belgrade, Military Academy, Veljka Lukića Kurjaka 33, Belgrade, milan.kresojevic@va.mod.gov.rs ORCID: 0000-0003-0278-5916

VESNA RISTIĆ VAKANJAC
University of Belgrade, Faculty of Mining and Geology, and full member of IRASA, Đušina 7, Belgrade, vesna.ristic@rgf.bg.ac.rs, ORCID: 0000-0002-2040-8252

DUŠAN POLOMČIĆ
University of Belgrade, Faculty of Mining and Geology, and full member of IRASA, Đušina 7, Belgrade, dusan.polomcic@rgf.bg.ac.rs, ORCID: 0000-0002-7807-1285

ALEKSANDAR TANASKOVIĆ
University of Belgrade, Faculty of Mining and Geology, Đušina 7, Belgrade, aleksandar.tanaskovic@rgf.bg.ac.rs, ORCID: 0009-0005-2668-3588

BORIS VAKANJAC
Metropolitan University, Faculty of Applied Ecology Futura, and full member of IRASA, Požeška 83a, Belgrade, Serbia, borivac@gmail.com, ORCID 0000-0002-9429-6943

JUGOSLAV NIKOLIC
Republic Hydrometeorological Service of Serbia, Knez Višeslava 66, Belgrade, jugoslav.nikolic@hidmet.gov.rs

Abstract: *The Velika Morava represents the largest river basin in the Republic of Serbia, with a total area of 37,444 km², of which 36,163 km² belongs to Serbia, accounting for 42.38 % of its total territory. In its lower course, it is a typical lowland river that frequently changes the geometry of its riverbed under the influence of natural and anthropogenic factors, which poses a major problem for the establishment of river navigation. River navigation on this river was established at the beginning of the 19th century, when the first steamboat arrived in Stalać, and during that period it was used exclusively for tourism purposes. Today, only the last 15 km of its course are navigable for the vessels of the River Flotilla in the execution of assigned tasks. However, the river flow itself and its hydraulic connection with groundwater formed in the alluvium of this river are of exceptional importance for the water supply of the population and industry, for agricultural purposes (irrigation), as well as for the extraction of sand and gravel from the riverbanks and the riverbed itself. The paper will present an analysis of natural and anthropogenic factors using statistical methods, as well as an assessment of their impact on the geometry of the river course, the regime of surface and groundwater, and, in this regard, the utilization of this natural resource for various purposes.*

Keywords: *water regime, river navigation, water supply, sand and gravel exploitation, Velika Morava*

1. INTRODUCTION

From the very beginning to the present day, human civilization and settlements have been tied to riverbanks and river courses themselves. The river was their source of everything needed for survival. Rivers provided sufficient amounts of water for water supply, irrigation of crops, and livestock watering. They also served as routes for river transport, sources of high-quality food (fishing), renewable sources of building materials (clay, gravel, and sand), and the like. However, it was also necessary to defend against rivers. This was particularly crucial during high water levels and the onset of flood waves. For this reason, historical records were initially kept indicating which areas the river flooded (the inundation area). Buildings were relocated and constructed beyond the influence of recorded flood waves, while digging wells in areas with settlements allowed the population to have accessible water for their own needs right where their living quarters were located.

Later, monitoring was established—initially for surface waters and subsequently for groundwater. One of the first instruments for tracking changes in river levels was the nilometer, used by the ancient Egyptians to monitor the Nile River. Much later, official services and institutions began to form. They established observation networks and conducted observations and measurements, initially of quantitative parameters (water levels and discharges) and later of qualitative parameters when that became necessary. By establishing a sufficiently long observation

Corresponding author: M. Kresojević;
E-mail: milan.kresojevic@va.mod.gov.rs

DOI: 10.5937/OTEH26154K

series of a quantitative parameter, it became possible to discern changes in their regimes resulting from both natural factors (climate change) and anthropogenic factors (dam construction and the formation of surface reservoirs, gravel and sand extraction, riverbed relocation for infrastructure projects like roads, or for expanding the extraction areas of existing or new mines, etc.). Many authors have analyzed the impact of natural and anthropogenic factors on changes in surface water regimes in their papers. For example, Wrzesiński & Sobkowiak [1] conducted a detailed analysis of the impact of natural and anthropogenic factors on changes in the water regime of the Vistula River (Poland), aiming to establish better water resource management and hydrological forecasting. A group of Ethiopian co-authors [2] addressed a similar topic, using the regime of the Gumara River (Ethiopia) as an example. Rempel & Church [3], focusing on the Fraser River (Canada), presented the impact of anthropogenic factors (gravel extraction) on the river course itself, as well as its overall environmental impact.

Since the banks and beds of larger rivers contain significant (mostly renewable) reserves of sand and gravel, extraction fields used for harvesting these building materials are often located on their banks. This paper presents how illegal and uncontrolled sand and gravel extraction from the banks of the Velika Morava River near the Ljubičevo Bridge affected the water resource of the Velika Morava itself, as well as the groundwater within the river's alluvium. During the 1990s, in the section near the Ljubičevo Bridge, this extraction led to the deepening of the Velika Morava riverbed by about 4 m, which reduced groundwater reserves in this area and degraded their quality. As a consequence, due to increased nitrate concentrations in the groundwater, the "Ključ" water source had to be disconnected from Požarevac's water supply system in 2006 [4], while the Trnovče water source (supplying Velika Plana and Smederevska Palanka) operates at a reduced extraction capacity [5].

In terms of basin size (37,966 km²) and length of its course, the Velika Morava is one of our largest rivers. Together with the Južna and Zapadna Morava, it drains just under 50% of Serbia's territory. The maximum discharge of the Velika Morava is 2,350 m³/s, the lowest recorded discharge is 25 m³/s, while the mean annual discharge stands at 257 m³/s [5]. This is a typical lowland river, featuring numerous meanders, oxbow lakes, and abandoned channels. From an agricultural production perspective, the alluvial plain of the Velika Morava represents high-quality soil suitable for cultivating grains, vegetables, industrial crops, and more. This has also influenced population density, though settlements are mostly located along the edges and less frequently on the alluvial plain itself.

On the other hand, river deposits of sand and gravel in the Velika Morava alluvium are among the most important construction materials and carry the character and significance of a natural resource. In addition to illegal gravel and sand extraction from the riverbed itself or its banks, it is a frequent occurrence for agricultural land to be purchased from owners, after which extraction begins on that very plot. The topsoil is mostly sold as high-quality earth, while the deeper layers are excavated for gravel and sand. This extraction has a major negative impact on neighboring agricultural lands, which are also quickly abandoned as extraction progresses, ultimately falling into the hands of those involved in gravel and sand extraction. This results in the permanent devastation of prime agricultural land [5].

Regarding navigability and river traffic, the Velika Morava was navigable along almost its entire length in the 19th century, from its confluence with the Danube to Stalać (the steamboat "Morava" sailed to Ćuprija and Stalać in 1867) [6]. Today, the Velika Morava is not navigable due to the dimensions of the waterway—primarily the depth, which has changed compared to the 19th century. This is the result of intensive erosion along its constituent rivers and main tributaries, primarily the Južna Morava, as well as the deposition of the mobilized material in the downstream sections of the Velika Morava. The aforementioned gravel and sand extraction, which led to the 4 m deepening of the riverbed downstream near the Ljubičevo Bridge, also caused an increase in the hydraulic gradient of the water table. Increased gradients drive more pronounced headward (regressive) erosion in certain upstream sections of the Velika Morava, causing further riverbed deepening upstream. Consequently, the Velika Morava is navigable for naval fleet vessels for only 15 km upstream from its confluence into the Danube River, as established by the Decree on International and Interstate Waterways [7].

2. METHODOLOGY

For the purpose of researching the impact of natural and anthropogenic factors (gravel and sand extraction from the banks and the riverbed itself) on the regime of surface and groundwater, all historical data on the water table elevations of the Velika Morava River at the Ljubičevo Bridge profile were collected, covering the period since monitoring was established (1923). Additionally, data on groundwater levels observed in piezometers placed perpendicular to the Velika Morava riverbed were gathered. Monitoring of the Velika Morava water levels was conducted daily (from 1923 to 1960 using a staff gauge, and from 1960 to the present day using a water level recorder/limnigraph). The groundwater regime is monitored using water level meters every 5 or 10 days. Groundwater monitoring in this section was established in 1977 and 1978. Subsequently, in 1987 and 1988, the existing network of observation facilities (piezometers) was densified, while during 2002, automatic water level loggers were installed for continuous groundwater monitoring. Figures 1 and 2 present comparative diagrams of the absolute water level elevations of the Velika Morava at the Ljubičevo Bridge gauging station and the groundwater level elevations recorded in the accompanying piezometers, covering the observation period from 1977 to 2018 (the period from the inception of groundwater monitoring up to 2018). Figure 1 refers to the right bank, and Figure 2 to the left bank [8].

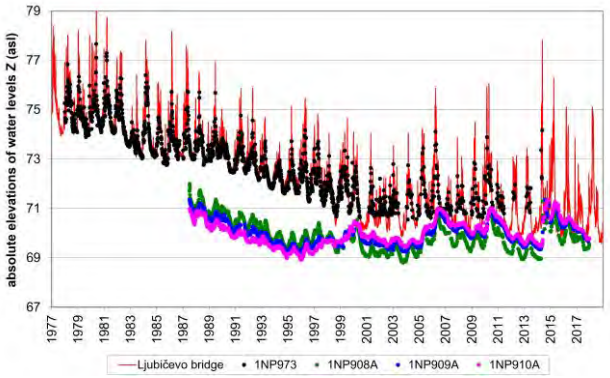


Figure 1. Comparative diagrams of the absolute water level elevations of the Velika Morava at the Ljubičevo Bridge gauging station and the groundwater levels observed in piezometers, observation period from 1977 to 2018 (right bank) [8]

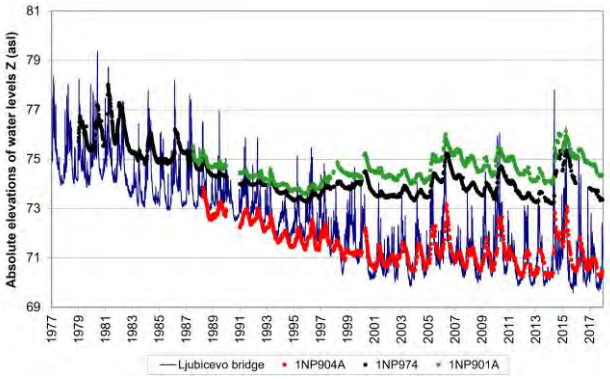


Figure 2. Comparative diagrams of the absolute water level elevations of the Velika Morava at the Ljubičevo Bridge gauging station and the groundwater levels observed in piezometers, observation period from 1977 to 2018 (left bank) [8]

It is clearly visible from diagrams 1 and 2 that the deepening of the Velika Morava riverbed affected both the water table elevations of the Morava and the groundwater levels. What is clearly evident from both Figure 1 and Figure 2 is that in piezometers 1NP973 and 1NP904A, which are located at the shortest distances from the Velika Morava riverbed itself, during the period from 1977 to 2000, the groundwater elevations were generally hypsometrically lower throughout the entire year; meaning that during this period, groundwater recharge occurred at the expense of the Velika Morava waters. After 2000, it is clearly visible that during low water levels, as well as during medium water levels, the groundwater elevations are at a higher elevation than the water levels of the Velika Morava. In other words, a change in the direction of groundwater flow occurred, so during these periods, the river is recharged at the expense of the groundwater.

To determine whether the aforementioned deepening of the riverbed affected the homogeneity of the water level series, homogeneity tests were applied. These tests showed that the examined samples extracted for the considered periods (before the riverbed deepening - the period before 1990, and after - the period after 1990) of the Velika Morava River level elevations observed at the Ljubičevo Bridge profile do not belong to the same population. More precisely, the extracted series are not homogeneous [9].

For the purpose of determining the homogeneity of the total discharge series, Student's t-test was also utilized, in which the criterion for evaluating the mean values is obtained using the following expression [10]:

$$t = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_1 \cdot \sigma_{X_1}^2 + n_2 \cdot \sigma_{X_2}^2}} \quad (1)$$

where: n_1 and n_2 are the sample sizes of time series 1 and 2, $\overline{X}_1$ and $\overline{X}_2$ are the mean values of time series 1 and 2, σ_{X_1} and σ_{X_2} are the standard deviations of time series 1 and 2.

The variable t follows a Student's distribution with $n_1 + n_2 - 2$ degrees of freedom. In this regard, the null hypothesis is accepted—meaning the series is homogeneous—if:

$$t_{\frac{\alpha}{2}} < t < t_{1-\frac{\alpha}{2}} \quad (2)$$

where α is the significance level, and in this specific case, a value of $\alpha = 0.05$ was chosen.

Furthermore, the paper places an emphasis on separating direct runoff and baseflow, with the aim of determining whether a change has occurred in any of the components of total runoff.

Baseflow is formed at the expense of groundwater located within the basin. The intensity of groundwater recharge decreases over time depending on the basin characteristics, geological composition of the terrain, and hydrogeological conditions—namely, the types of aquifers. This is of great importance because the terrain's capacity to retain water that has fallen on the basin underground, and subsequently return it to the river course, represents one of the more crucial basin characteristics, which largely defines the runoff regime during dry periods. Rivers whose basins are primarily composed of fractured aquifers or provisionally arid terrains frequently dry up during prolonged dry periods. Therefore, during these periods, the aquifer exhausts all of its dynamic reserves on recharging the surface course. After that, the river channel remains dry because the aquifer runs out of dynamic reserves. Consequently, during low and mostly medium water levels, the water in the riverbed is sustained by groundwater.

Since the observed (total) discharge consists of direct runoff—which is a consequence of precipitation—and baseflow—which is formed through the recharge of the surface course by groundwater—this paper first performed the separation of direct runoff from baseflow. For this purpose, the BFI+ software [11] was applied, utilizing the Local Minimum method [12]. This method identifies days with the lowest flow within a defined time interval—half of the observed period minus one day before and after the analyzed day. If the flow on that day

is the lowest, it is designated as a local minimum and connected to neighboring minima using straight lines. The baseflow between these points is estimated via linear interpolation, thereby connecting the lowest values on the hydrograph into a simplified curve [13].

After calculating the direct runoff and baseflow and applying the homogeneity test to the formed series, the Baseflow Index (BFI) was also calculated. This index represents a quantitative indicator of the share of baseflow in the total river discharge at a specific profile. Essentially, the BFI represents the ratio between the volume of baseflow, V_B and the total volume of discharged water, V_T :

$$BFI = \frac{V_B}{V_{uk}} \tag{3}$$

where:

$$V_T = \overline{Q} \cdot T \tag{4}$$

and where $\overline{Q}$ represents the mean multi-year discharge, while T represents a one-year period expressed in seconds.

Depending on the geological composition of the terrain, namely the capacity of the subsurface to accumulate and release water, the BFI value varies within very wide limits, ranging from 0.15 to 0.90 [10].

3. RESULTS

By applying the aforementioned methodology, baseflow and direct runoff were calculated. Figures 3 and 4 present the hydrographs of total and baseflow runoff, specifically in Figure 3 for the first 10 years of the calculated series (1948–1957) and in Figure 4 for the last 10 years of the observation series (2009–2018). The hydrograph characterizing the entire observation period is not shown in the figure due to the length of the considered timeframe.

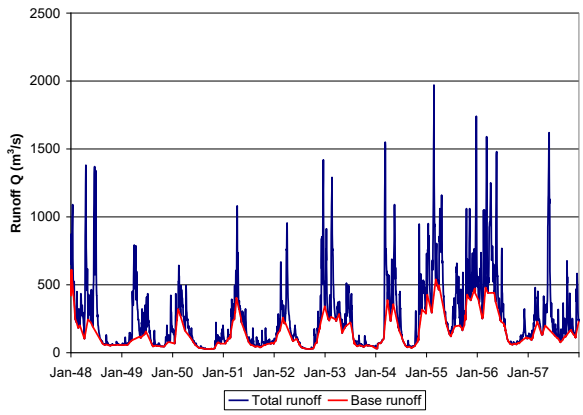


Figure 3. Hydrograph of total and baseflow runoff of the Velika Morava, Ljubičevo Bridge gauging station, period 1948–1957

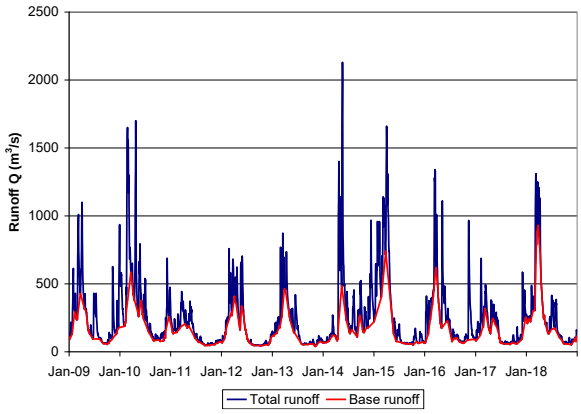


Figure 4. Hydrograph of total and baseflow (runoff) of the Velika Morava, Ljubičevo Bridge gauging station, period 2009–2018

After calculating the baseflow and direct runoff, and with the aim of applying the homogeneity test to the obtained series (total, direct, and baseflow runoff), Table 1 presents the statistics (mean value and standard deviation) for the separated periods (the first period 1948–1989 and the second period 1990–2018). Table 2 presents the calculated statistics as well as the conditions for accepting the null hypothesis.

Table 1. Mean values and standard deviations of total, direct, and baseflow runoff (m³/s)

	Q_t		Q_b		Q_d	
	I period	II period	I period	II period	I period	II period
Q_{av}	244.7	221.1	158.4	153.8	86.28	67.32
σ	70.511	72.813	43.056	45.940	35.738	32.186

Table 2. Calculated t-statistics and conditions for accepting the null hypothesis, i.e., the hypothesis that the series are homogeneous

	t	$t_{\alpha/2}$	$t_{1-\alpha/2}$	homogeneity
Q_t	1.346	-2.00	2.00	yes
Q_b	0.425	-2.00	2.00	yes
Q_d	2.254	-2.00	2.00	no

Based on the obtained values and t-statistics, it can be concluded that the 4 m deepening of the Velika Morava riverbed did not have a statistically significant impact on the change in total and baseflow runoff, while a certain change appeared in the direct runoff. In addition to the anthropogenic impact, climate change—specifically the increasing frequency of intense precipitation or its absence—certainly had a certain influence on such a result.

Furthermore, based on the obtained baseflow values, the base flow index (BFI) was calculated using equation (3). Figure 5 presents the diagram of the BFI change over time. The same diagram also displays the trend line, as well as the mean BFI values for the period before 1990 and after 1990.

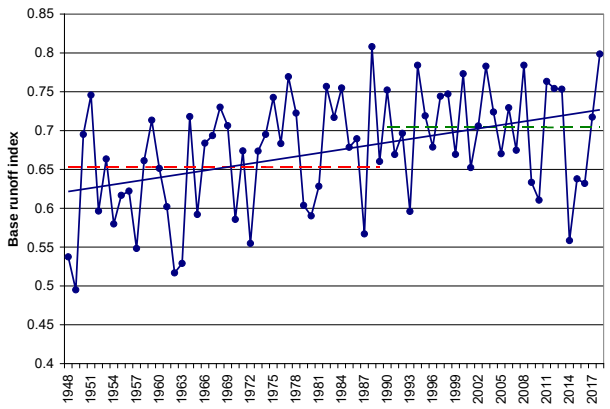


Figure 5. Diagram of the base flow (runoff) index change in the period from 1948 to 2018, featuring the trend line and multi-year mean values calculated for the separated periods (before and after 1990)

4. DISCUSION WITH CONCLUSION

Based on Diagram 5, it can be concluded that, on average, the base flow index values in the period from 1948 to 1989 are 0.05 lower than those characteristic of the period from 1990 to 2018. The BFI trend is positive, and its multi-year mean value for the period before 1990 is 0.653, while for the period from 1990 to 2018, it stands at 0.704. The lowest BFI values are associated with exceptionally rainy years, when direct runoff, formed as a result of intense precipitation, plays a more significant role in the Velika Morava riverbed. In contrast to rainy years, the BFI reaches its highest values during dry periods. In the first period, the absolute minimum BFI was calculated for the year 1949, amounting to 0.495, while the absolute maximum was recorded in 1988, reaching 0.808. During the period after 1990, an absolute minimum BFI of 0.558 was obtained for 2014—the year when record-breaking May rainfall caused by Cyclone Tamara led to severe flooding in western and central Serbia. The absolute maximum (BFI = 0.798) was recorded in 2018.

Based on the presented data, it can be concluded that there has been an increase in the proportion of groundwater forming the total runoff of the Velika Morava. This is certainly partly a consequence of the change in the direction of groundwater flow during low and medium water periods; meaning that for most of the year now, the first aquifer discharges to recharge the Morava River. The surface watercourse recharges the aquifer only during high water periods.

The lowering of the Morava River level due to the deepening of its riverbed (as previously discussed) consequently led to a drop in the groundwater levels of the first aquifer formed within the river's alluvium. Together with the change in the direction of groundwater flow—namely, the depletion of groundwater reserves to recharge the Morava River—these factors have driven the significant reduction in groundwater reserves currently observed. According to some calculations, based on an average groundwater level decline of about 1.5 m in the alluvial plain area of the Velika Morava (within the section belonging to the sheet of the Basic Hydrogeological Map of Požarevac, 1:100,000, and relative to the 1975–1985 period), the calculated volume of "lost" groundwater amounts to approximately 150.5 x 10^6 m³ (16.5%), out of the total estimated permanent reserves of 756 x 10^6 m³ of water [14]. As a result of diminished groundwater reserves, their quality degraded. Specifically, due to the use of artificial and natural fertilizers (manure), combined with the reduced groundwater reserves, nitrate concentrations exceeded the maximum allowable limit of 50 mg/dm³ after the 1990s. For this reason, for example, the "Ključ" water source, whose waters are used for the water supply of Požarevac, was taken out of service in 2006. The problem in this area was resolved by constructing infiltration basins, which helped lower the concentration of nitrate ions in the water source zone. Thus, in addition to reducing groundwater quantities, the deepening of the Velika Morava riverbed near the Ljubičevo Bridge consequently impacted the quality of these waters as well.

In addition to the aforementioned issues, there is also the problem of intense riverbed clogging (clogging/clogging of the riverbed) in certain sections, caused by an increased content of suspended material resulting from sand and gravel extraction. The result of riverbed clogging is a lowering of the water level in the alluvial aquifer due to a reduction in the contact area between the aquifer and the riverbed. This directly results in a decrease in the volume of water that infiltrates into the subsurface to be used for public water supply. The ultimate outcome is, once again, reduced groundwater reserves in this area.

Because of this, in the area of the "Trnovče" regional water source, whose waters are used for the water supply of Velika Plana and Smederevska Palanka, highly yield-abundant groundwater localities are not located in the immediate vicinity of the riverbed, but rather in the zone of a fossil river channel (meander) that has remained in direct hydraulic connection with the Velika Morava River. Thus, at a distance of about 100 m from the riverbank, vertical water intakes can capture maximum groundwater quantities of only Q = 3–6 l/s, whereas moving deeper toward the hinterland, wells of the same type can yield quantities of Q = 10–25 l/s [15].

The lowering of groundwater levels has caused many shallow dug wells, whose waters were used for irrigating agricultural crops, to run dry during arid periods (when irrigation is most necessary). Consequently, problems with irrigating agricultural areas have also emerged.

Another highly critical impact is on the navigability of the river itself. The natural course of the Morava River in its lower section is meandering in character; therefore, the riverbed shifts through natural processes, which in this area are now also driven by the extraction of construction materials from the banks and the riverbed itself. This extraction affects the displacement of the thalweg (the main current line), which, for example, can shift from the central section toward one of the banks where extraction is taking place, thereby accelerating bank erosion. In this manner, riverbanks collapse, the river meanders more intensely, and the bed depth changes frequently, which greatly complicates the establishment of navigability on

the Velika Morava.

However, by expanding its waterways, the Republic of Serbia could potentially realize major economic benefits, primarily by increasing the transport corridor of the cheapest and most environmentally sustainable mode of transport within its territory. Integrating the Velika Morava into the system of navigable rivers in the Republic of Serbia would certainly be a massive undertaking. Furthermore, it would be of immeasurable importance from the perspective of national defense, both regarding military threats and in the event of natural disasters. Expanding the waterway network would provide River Flotilla (RF) units with an increased scope for operational deployment, while a potential enemy advance would be hindered. From a national defense standpoint, a dense network of waterways and canals intersecting flat, easily accessible terrains impedes the operational advancement of enemy forces.

We must remain cognizant of these facts in order to change the irresponsible attitude toward water resources, and to reduce gravel extraction within the riparian zone and bed of the Velika Morava River to a reasonable level—one that will not negatively impact the development of existing water sources, all with the ultimate goal of improving public water supply in the future.

References

[1] WRYESIŃSKI,D., SOBKOWIAK,L.: *Transformation of the Flow Regime of a Large Allochthonous River in Central Europe—an Example of the Vistula River in Poland*. Water, 12, 507 (2020).

[2] ABEBE,W.B., TILAHUN,S.A., MOGES,M.S., WONDIE,A., DERSEH,M.G., NIGATU,T.A., MHIRET,D.A., STEENHUIS,T.S., CAMP,M.V., WALRAEVENS,K., McCLAIN,M.E.: *Hydrological Foundation as a Basis for a Holistic Environmental Flow Assessment of Tropical Highland Rivers in Ethiopia*. Water, 12, 547 (2020)

[3] REMPEL,L.L., CHURCH,M.: *Physical and ecological response to disturbance by gravel mining in a large alluvial river*. Can. J. Fish. Aquat. Sci., 66, 52–71 (2009)

[4] STEPANOVIĆ,M., BAJIĆ,D., POLOMČIĆ,D., AVRAMOVIĆ,A., MIJATOVIĆ,B., *Kvalitativne karakteristike podzemnih voda izvorišta „Ključ“ u Požarevcu*, XVI, Srpski hidrogeološki simpozijum sa međunarodnim učešćem, pp 495-500, Beograd (2022)

[5] POPOVIĆ,LJ., POPOVIĆ,Z., RISTIĆ VAKANJAC,V., VAKANJAC,B., MANDIĆ,M.: *Trenutno stanje i moguće posledice nekontrolisane eksploatacije šljunka po životnu sredinu i vodosnabdevanje - primer Velike Morave*, XV Srpski simpozijum o hidrogeologiji sa međunarodnim učešćem, Rudarsko-geološki fakultet, Univerzitet u Beogradu, pp. 435-440, Kopaonik, Srbija, 2016

[6] KRESOJEVIĆ,M., RISTIĆ VAKANJAC,V., TRIFKOVIĆ,D.: *Proširivanje mreže plovnih puteva Republike Srbije u funkciji odbrane zemlje*, SymOpIs 23, Tara, pp 727-732, 2023

[7] VLADA REPUBLIKE SRBIJE, *Uredba o određivanju međunarodnih i međudržavnih puteva*, Službeni glasnik Republiek Srbije (109/2016 i 68/2019), Beograd

[8] RISTIĆ VAKANJAC,V., KRESOJEVIĆ,M., POLOMČIĆ,D., VAKANJAC,B., NIKOLIĆ,J.: *Uticaj eksploatacije šljunka i peska na režim površinskih i podzemnih voda*, Časopis "Voda i sanitarna tehnika" LV(3) 15-24 (2025) (ed. Jasmina Agbaba) ISSN 0350-5049

[9] KRESOJEVIĆ,M., RISTIĆ VAKANJAC,V., TRIFKOVIĆ,D., NIKOLIĆ,J., VAKANJAC,B., POLOMČIĆ,D., BAJIĆ,D., *The effect of gravel and sand mining on groundwater and surface water regimes – a case study of the Velika Morava river, Serbia*, *Water* **2023**, *15*(14)

[10] PROHASKA,S.: *Hidrologija - I deo*. Rudarsko-geološki fakultet, Univerzitet u Beogradu, Beograd. 2004

[11] HYDRO OFFICE: *Software for hydrology, meteorology, geoscience & data science*. – Available from: https://hydrooffice.org., 2020

[12] WAHL,K.L., WAHL,T.L.: *Determining the flow of Comal Springs at New Braunfels, Texas.* – In: Proc. Texas Water ’95, component conference of the First Int. Conf. on Water Resources Engineering. San Antonio, American Society of Civil Engineers, 77–86., 1995

[13] SLOTO,R.A., CROUSE,M.Y.: *HYSEP: A computer program for streamflow hydrograph separation and analysis*. – U.S. Geological Survey, Water-Resources Investigations Report 96-4040, Pennsylvania, 46 p, 1996

[14] MARKOVIĆ,D. i dr. 2014: Tumač-a za Osnovnu hidrogeološku kartu Republike Srbije, list Požarevac 1: 100 000, Geološki zavod Srbije, Beograd (Revidovana verzija)

[15] POPOVIĆ,Lj.: *Izveštaj o izradi eksploatacionih bunara: BT-6, 8, 9, 10, 11, 12 na izvorištu "Trnovče" kod Velike Plane*, Fond: Geološki institut Srbije, Beograd, 2006

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

COMPARATIVE ANALYSIS OF DIFFERENT SYSTEMS FOR PROJECTILE VELOCITY MEASUREMENT IN BALLISTIC TESTING

ALEKSA ANIČIĆ
Agency for Testing, Stamping and Marking of Weapons, Devices and Ammunition, Kragujevac, tehnika@aiz.rs

NEBOJŠA HRISTOV
Military Academy, University of Defence, Belgrade, nebojsahristov@gmail.com

LAZAR BLAGOJEVIĆ
Agency for Testing, Stamping and Marking of Weapons, Devices and Ammunition, Kragujevac, proofhouse@aiz.rs

MILOŠ JONČIĆ
Ministry of Foreign Affairs, Arms Control Department, Belgrade, milos.joncic@mfa.rs

ILIJA LJUJIĆ
Zastava Arms, Quality Control Department, Kragujevac, kkvaliteta@zastava-arms.rs

DAMIR JERKOVIĆ
Military Academy, University of Defence, Belgrade, damir.jerkovic@va.mod.gov.rs

Abstract: *The paper presents the comparative analysis of projectile velocity measurement results obtained using different types of measuring equipment in ballistic testing. The considered velocity measurement systems include infrared barriers and various types of radar systems, with the aim of assessing their applicability, accuracy, reliability, and limitations under different test conditions. In addition to the comparative analysis of experimentally obtained results, the paper also provides an overview of different types of measuring equipment, their structural design, operating principles, and basic technical characteristics relevant to their application in ballistic testing. The tests were carried out using a ballistic barrel receiver, a test barrel, and accompanying measuring equipment, for several different calibers within the category of small-arms ammunition. The comparison of the measured projectile velocity values obtained by different measuring systems was performed, together with the analysis of the influence of the type of measuring equipment, measuring position, test configuration, and test conditions on the obtained results. Particular attention was given to the analysis of the estimation of measurement uncertainty in order to evaluate the validity and repeatability of the obtained data. Based on the obtained results, the assessment of the applicability of different measuring systems in ballistic testing of small arms was given.*

Keywords: *Projectile Velocity, Ballistic Testing, Measuring Equipment, Ammunition.*

1. INTRODUCTION

In the field of ballistics, measuring projectile velocity is one of the most important tasks. Projectile velocity directly determines kinetic energy, stability, and terminal effectiveness. Without precise velocity data, it is impossible to develop modern weapons, control the quality of gunpowder and ammunition, or test the resistance of ballistic protective equipment.

In the past, mechanical and electrical methods, such as the ballistic pendulum or wire grids, were used to measure projectile velocity. With older systems, the projectile had to physically hit the pendulum or cut wires at a precisely defined distance to calculate the projectile's velocity. The procedure was complicated to set up, required constant replacement of destroyed wires, and often yielded inaccurate results. Today, technology has completely changed this process and shifted to non-contact measurement. Modern devices allow the projectile to pass through the measurement point without any physical contact, while electronic measurement stations instantaneously read data on the screen with extremely high precision [5, 6, 8, 9].

In this study, two such modern but technologically completely different measuring systems were used: a ballistic chronograph based on infrared barriers *Top Trend V2.205* (Figure 1), and a portable Doppler radar *Garmin Xero C1 Pro* (Figure 2). The ballistic chronograph based on IR barriers (hereinafter referred to as the optical chronograph) operates on an optical principle – it measures the time required for a projectile to pass between two invisible IR light beams [3, 6, 9]. On the other hand, the Garmin (hereinafter referred to as the radar chronograph) utilizes radar technology, where the device emits radio waves that bounce off the flying

Corresponding author: A. Aničić;
E-mail: tehnika@aiz.rs

DOI: 10.5937/OTEH26155A

projectile, thereby continuously tracking its velocity.

To compare the performance of the aforementioned measuring systems, three calibers were selected for this research, covering a wide range of applications.

Figure 1 Ballistic chronograph based on infrared barriers, Top Trend V2.205 (optical chronograph)

Figure 2 Garmin Xero C1 Pro

The calibers used during the ballistic testing are as follows:

- 9 mm Luger (Figure 3a)
- 7.62x39 mm (Figure 3b)
- .308 Winchester (7.62x51 mm) (Figure 3c)

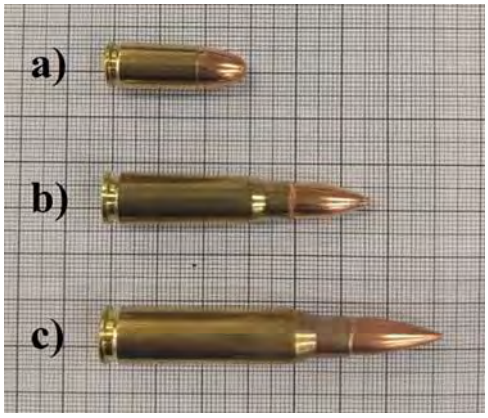


Figure 3 Ammunition: a) 9 mm Luger; b) 7.62x39 mm; c) .308 Winchester (7.62x51 mm)

Figure 4 shows the test barrels used for the testing.

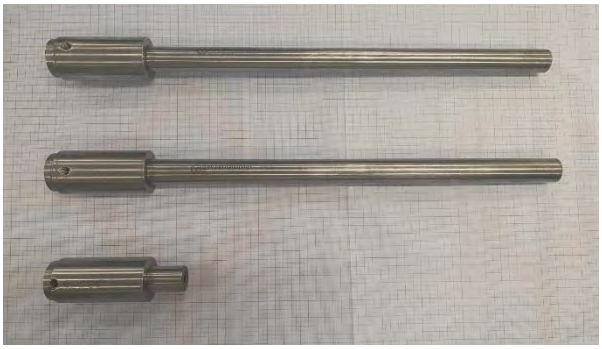


Figure 4 Test barrels (in order): 7.62x39 mm; .308 Winchester; 9 mm Luger

This study aims to perform a comparative analysis of the measuring systems through experimental testing. Through statistical processing of the measured velocity values for the specified three calibers, the paper aims to determine the level of measurement uncertainty for both systems, identify their operational advantages and limitations, and provide recommendations for their optimal application in research, development, and testing ballistic processes.

2. EXPERIMENTAL PERFORMANCE TESTING

This chapter details the method of executing the test, the spatial layout of the measuring components in the ballistic tunnel, as well as the basic physical principles on which the operation of the applied measuring systems is based.

For the purpose of measuring the muzzle projectile velocities of the selected calibers (9 mm Luger, 7.62x39 mm and .308 Winchester), specialized ballistic barrels were used. These barrels were manufactured, mounted, and calibrated in strict accordance with the international C.I.P. standard (*Commission internationale permanente pour l'épreuve des armes à feu portatives*). According to the C.I.P. standard, projectile velocity is measured at a distance of 2.5 m from the muzzle, with the chronograph positioned to determine the projectile velocity at that reference distance [1].

The use of standardized barrels in ballistic testing is crucial because they eliminate the influence of mechanical tolerances, wear and gas losses characteristic of service weapons. This ensures controlled, maximally repeatable initial conditions and allows the generated pressures and projectile velocities to be solely a function of the propellant charge characteristics and the geometry of the cartridge itself.

Ammunition conditioning was also performed in a temperature control chamber to prevent external environmental factors from negatively affecting the test. The temperature inside the chamber was 21°C, while the air humidity was 65%. The ammunition was stored in the chamber for 24 hours prior to the test itself. Table 1 shows the characteristics of the tested ammunition.

Table 1 Characteristics of the ammunition used for testing

Caliber	Projectile type	Bullet mass (gr)	Projectile mass (g)	Nominal muzzle velocity (m/s)
9 mm Luger	FMJ	124	8,00	~340
7,62x39	FMJ	123	8,00	~715
.308 Win	FMJ BT	145	9,40	~850

Before starting the test, the measuring devices were placed in specific positions to ensure simultaneous recording of the velocity of the same projectile during a single firing. Figure 5 shows the spatial configuration of the measuring systems, with the optical chronograph positioned on the right side, while the radar chronograph is positioned on the left side. They are located at a distance of 2.5 m ± 5 cm from the muzzle.

Figure 5 Spatial configuration of measuring systems: infrared barrier (right) and Garmin Xero C1 Pro (left)

The applied measuring devices represent two fundamentally different technological solutions in the field of experimental kinematics. The optical chronograph (Figure 1) functions as a passive-active optoelectronic discrete system for measuring passage time. Its operation is based on forming a continuous light curtain in the infrared spectrum within a rigid metal frame. When an opaque body (projectile) cuts through the first light plane, an instantaneous change in light intensity occurs at the photodetectors, which electronically activates a high-resolution time counter (timer). Passage through the second light plane, located on a factory-calibrated measuring base, generates a second pulse that stops the counting. The projectile velocity is subsequently interpreted as the average velocity over that fixed distance [3, 6, 9]. Table 2 provides the technical specifications of the optical chronograph.

Table 2 Technical specifications of ballistic chronograph based on infrared barriers, *Top Trend V2.205[3]*

Effective measuring area	1000x1000 mm
Caliber	4 – 40 mm
Velocity measurement range	30 – 3000 m/s
Fixed mesurement base	2000 mm
Measurement error	±0,1%
Spectral range	790 – 1050 mm (IC – A)
Detection maximum	940 nm
Output signal: square wave pulse	≤5V DC
Operating range	-10°C do +50°C
Power supply	230 V AC, 50 – 60 Hz
Weight	35 kg

The radar chronograph (Figure 2) functions as an active continuous-wave high-frequency radar system. The device continuously emits electromagnetic waves of a defined frequency in the direction of the projectile's movement. When the emitted wave encounters a projectile in flight, it reflects and returns to the device's receiving antenna. Given that the projectile moves at a velocity significant relative to the wave propagation velocity, the frequency of the reflected signal shifts proportionally to the body's velocity (Doppler shift). The system performs real-time frequency analysis (FFT) of the received signal, transforming the measured frequency shift directly into the instantaneous velocity of the projectile [2, 5, 8]. Table 3 provides the technical specifications of this measuring system.

Table 3 Technical specifications of the Garmin Xero C1 Pro [2]

Measuring range	30 do 1600 m/s
Dimension	77 x 60,5 x 34,44 mm
Weight	105,6 g
Water rating	IPX7
Display type	Sunlight-visible, monochrome
Display size	133 x 44 mm, diagonal 55 mm
Battery type	Internal rechargeable lithium-ion battery; non replaceable

Battery life	Up to 2000 shots or up to 6 hours
Operating range	-10°C to +55°C (14°F to 131°F)
Memory/history	50 sessions at max 100 shots per round before a Xero Core app sync is required to save data to phone
Interface	USB-C

3. RESULTS AND DISCUSSION

After setting up the measurement configuration, testing was conducted for each caliber individually. For each caliber, 20 shots were fired in a single group, during which the results provided by the optical chronograph were read alongside the muzzle velocity value measured by the radar chronograph. The tables presented below, in addition to the muzzle velocity values of each individual projectile, contain the average measured velocity value, maximum velocity, minimum velocity, spread of values (extreme spread) and standard deviation.

The first test was conducted for the 9 mm Luger caliber, and the obtained values read from the measuring systems are compared in Table 4.

Table 4 Muzzle velocity of the 9 mm Luger projectile measured by two different measuring systems

9 mm Luger		
No.	Optical chronograph	Radar chronograph
1.	387,60	388,40
2.	391,54	392,80
3.	381,97	380,40
4.	388,73	386,10
5.	384,76	383,00
6.	384,17	384,20
7.	386,03	386,60
8.	387,22	387,90
9.	383,29	384,60
10.	375,38	377,00
11.	377,71	376,00
12.	387,37	387,90
13.	381,61	380,40
14.	389,11	390,30
15.	380,01	381,00
16.	381,90	379,20
17.	381,61	382,60
18.	388,50	390,70
19.	385,73	387,80
20.	373,13	373,80
Avg.val.	383,87	384,04
Max:	391,54	392,80
Min:	373,13	373,80
Spread:	18,41	19,00
St. dev.	4,79	5,21

Based on the 20 performed measurements, it was determined that the average measured velocity for the optical chronograph is 383.87 m/s, while for the radar chronograph it is 384.03 m/s. The difference between the average values is only 0.16 m/s, indicating very good agreement between the results.

The spread of measured values for the optical chronograph is 18.41 m/s, and for the radar chronograph it is 19 m/s, which shows that both devices record almost identical velocity variations. The standard deviation for the optical chronograph is 4.79 m/s, while for the radar it is 5.21 m/s, implying low dispersion of results and good measurement repeatability for both systems.

The obtained results show that the radar chronograph yields almost identical results to those provided by the optical chronograph. The small difference in average values, as well as the similar values for spread and standard deviation, indicate that the radar chronograph provides reliable and stable measurements of the muzzle velocity for the 9 mm Luger projectile.

Following the 9 mm Luger caliber, the muzzle velocity measurement was performed for the 7.62x39 mm caliber. Table 5 presents the measurement results.

Table 5 Muzzle velocity of the 7.62x39 mm projectile measured by two different measuring systems

7,62x39 mm		
No.	Optical chronograph	Radar chronograph
1.	725,16	725,40
2.	736,38	737,10
3.	738,28	738,70
4.	737,46	736,80
5.	737,46	737,40
6.	735,56	735,50
7.	736,11	736,40
8.	738,55	738,30
9.	736,92	737,90
10.	731,26	731,70
11.	737,46	737,60
12.	736,11	736,70
13.	742,39	744,30
14.	737,74	737,90
15.	736,65	736,30
16.	739,64	740,60
17.	745,71	746,10
18.	741,02	741,30
19.	737,46	738,60
20.	737,19	737,50
Avg.val.	737,23	737,61
Max:	745,71	746,10
Min:	725,16	725,40
Spread:	20,55	20,70
St. dev.	4,04	4,20

Based on the 20 performed measurements, it was determined that the average measured velocity for the optical chronograph is 737.23 m/s, while for the radar it is 737.60 m/s. The difference between the average values is

only 0.38 m/s, indicating very good agreement between the results obtained by both measuring systems.

The spread of measured values for the optical chronograph is 20.55 m/s, and for the radar it is 20.70 m/s, which shows that both devices record almost identical velocity variations. The standard deviation for the optical chronograph is 4.04 m/s, while for the radar chronograph it is 4.20 m/s, implying low dispersion of results and good measurement repeatability for both systems.

The obtained results show that the radar chronograph yields almost identical results to those provided by the optical chronograph. The small difference in average values, as well as the similar values for spread and standard deviation, indicate that the radar chronograph provides reliable and stable measurements of the muzzle velocity for the 7.62x39 mm projectile.

The next measurement was performed for the .308 Winchester (7.62x51 mm) caliber.

Table 6 Muzzle velocity of the .308 Winchester projectile measured by two different measuring systems

.308 Winchester		
No.	Optical chronograph	Radar chronograph
1.	847,10	838,70
2.	858,00	866,60
3.	862,07	859,70
4.	853,24	852,80
5.	856,53	855,40
6.	868,43	867,30
7.	852,88	853,80
8.	862,07	862,00
9.	859,85	859,50
10.	867,30	866,70
11.	861,70	861,20
12.	867,30	871,80
13.	848,90	848,10
14.	853,97	855,50
15.	856,16	854,00
16.	862,44	861,80
17.	854,70	854,10
18.	862,44	861,80
19.	849,98	850,10
20.	857,63	855,70
Avg.val.	858,13	857,83
Max:	868,43	871,80
Min:	847,10	838,70
Spread:	21,33	33,10
St. dev.	6,15	7,65

Based on the 20 performed measurements, it was determined that the average measured velocity for the optical chronograph is 858.13 m/s, while for the radar chronograph it is 857.83 m/s. The difference between the average values is only 0.30 m/s, indicating very good agreement between the results obtained by applying both measuring systems.

The spread of measured values for the optical chronograph is 21.33 m/s, while for the radar chronograph it is 33.10 m/s. The standard deviation for the optical chronograph is 6.15 m/s, and for the radar it is 7.65 m/s, indicating less dispersion of results and slightly better measurement repeatability for the optical chronograph compared to the radar chronograph during this test.
Although a slightly larger spread of measured values and a larger standard deviation were recorded for the radar chronograph, the average values obtained by both devices are practically identical. Based on the obtained results, it can be concluded that the radar chronograph provides reliable and stable measurements of the muzzle velocity for the 7.62x51 mm projectile, with the results agreeing very well with those obtained using the optical chronograph.

4. RESULTS ANALYSIS

Following the testing and the obtained data, it is necessary to analyze the results, which is done in this chapter. An analysis of the results obtained from the measurements of the optical chronograph and the radar chronograph was conducted. The aim of the analysis is to compare the results obtained with the two measuring systems, evaluate their mutual agreement, and examine the repeatability and reliability of the measurements. The analysis is based on a comparison of average velocity values, spreads of measured values, standard deviations, percentage deviations, and measurement uncertainty for all three tested ammunition calibers.

4.1 Analysis of average values

Figure 5 shows a comparative diagram of the average velocity values for all three calibers (both measurement systems).

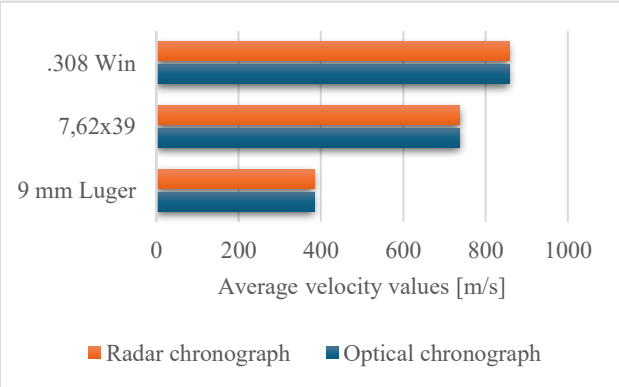


Figure 5 Comparative presentation of the obtained average velocity values using different measuring systems for three different calibers

The diagram clearly shows that the average velocity values of both measuring systems are almost identical, indicating consistency between the two systems.

4.2 Analysis of measurement repeatability

Figure 6 shows the comparative values of the standard deviation obtained by measuring with both systems for three different calibers.

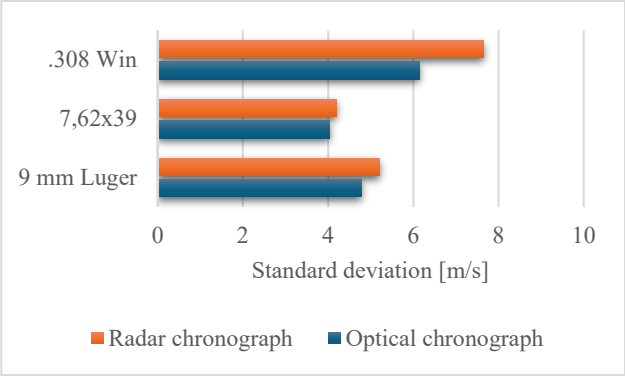


Figure 6 Comparative presentation of the obtained standard deviations

A smaller standard deviation value indicates higher measurement repeatability. For all three calibers, it was observed that the standard deviation obtained by measuring with the optical chronograph is smaller. The minimal difference between the two systems is noticed for the 7.62x39 caliber, while the largest difference between the standard deviation values was obtained during the testing of the .308 Winchester caliber.

4.3 Percentage deviation

To show the percentage deviation, the following formula was used:

$$\delta = \frac{\left|V_{avr(optical)} - V_{avr(radar)}\right|}{V_{avr(radar)}} * 100\%$$

Based on the tables from Chapter 3, the following percentage deviations were obtained:

- 9 mm Luger: 0.04%
- 7.62x39 mm: 0.05%
- .308 Winchester: 0.04%

The percentage deviations for all three calibers are approximately equal, indicating very small deviations between the measurement results.

4.4 Measurement uncertainty

Measurement uncertainty is calculated based on the obtained standard deviation[4, 7]. First, it is necessary to calculate the standard uncertainty:

$$u = \frac{Standard\ deviation\ value}{\sqrt{n}}$$

where *n* represents the total number of shots.

After calculating the standard uncertainty, the expanded uncertainty is calculated as follows:

$$U = 2u$$

Based on the aforementioned relations, the obtained results for the expanded measurement uncertainty are presented in the diagram in Figure 7.

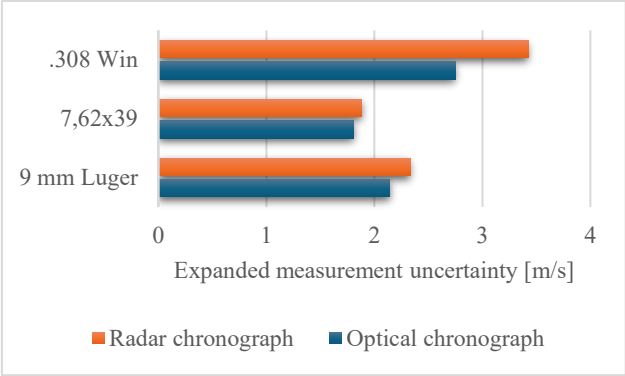


Figure 7 Comparative presentation of expanded measurement uncertainty

4.5 General analysis

Based on all performed measurements, it can be concluded that the optical and radar chronographs provide reliable, stable, and repeatable measurements of muzzle projectile velocity. The analysis of average values, spread, standard deviation, expanded measurement uncertainty, and percentage deviation showed a high degree of consistency between the results obtained by applying both ballistic chronographs for all three tested calibers. The observed differences are small and within the expected measurement uncertainty, indicating that the radar chronograph can be successfully used as an alternative to the optical chronograph for measuring muzzle projectile velocity [5, 8].

CONCLUSION

The aim of this study was to test the measuring systems to see if approximate results are obtained using two different systems. Tests were performed for three different ammunition calibers: 9 mm Luger, 7.62x39 mm, and .308 Winchester, with 20 individual measurements realized for each caliber.

An analysis of the obtained results determined that the average values of the measured velocities are very similar for both measuring systems. Also, the analysis of spread, standard deviation, expanded measurement uncertainty, and percentage deviation showed that the differences between the results are small and within the expected measurement uncertainty. The percentage deviations for all three calibers amounted to approximately 0.04-0.05%, indicating a high degree of agreement between the results.

The calculated values of spread, standard deviation, and measurement uncertainty show that both measuring systems have good measurement repeatability. The obtained results confirm that the optical and radar chronographs can be successfully used to measure initial projectile velocity in laboratory and testing conditions.

The results of this work represent a basis for further research, which may include a larger number of different ammunition types, testing under different environmental conditions, and the analysis of additional factors affecting measurement accuracy. Such research would enable an

even more complete evaluation of the measurement characteristics of the optical chronograph as well as the radar chronograph and their application in ballistic testing.

Based on all conducted tests, it can be concluded that the optical and radar chronographs represent reliable and practical solutions for measuring muzzle projectile velocity, and the obtained results confirm that their application can satisfy the requirements of experimental ballistic testing.

References

[1] C.I.P. – Permanent International Commission for the Proof of Small Arms: *Decisions, Texts and Tables: Procedures for Measurement of Projectile Velocity*, Brussels, Belgium, edition used in the testing laboratory,2026.

[2] Garmin Ltd.: Xero C1 Pro Chronograph – Owner's Manual and Technical Specifications, Garmin Ltd., Olathe, KS, USA, 2023.

[3] Top Trend: V2.205 Ballistic Chronograph Based on Infrared Barriers – Technical Documentation and Operating Instructions, manufacturer's technical documentation, Lučani, Serbia, 2020.

[4] Joint Committee for Guides in Metrology: *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement*, JCGM 100:2008, BIPM, 2008, doi:10.59161/JCGM100-2008E.

[5] Knudsen, P. J. T., Svender, J.: Doppler Radar Velocity Measurements for Wound Ballistics Experiments, *International Journal of Legal Medicine*, 1994, Vol.107, No.1, pp.1–6, doi:10.1007/BF01247266.

[6] Paulter, N. G. Jr., Larson, D. R.: Reference Ballistic Chronograph, *Optical Engineering*, 2009, Vol.48, No.4, Article 043602.

[7] Paulter, N. G. Jr.: Measurement Uncertainty of Bullet Velocities that Are Acquired with Light Screens and a Waveform Recorder, *Journal of Testing and Evaluation*, 2023, Vol.51, No.6, doi:10.1520/JTE20230061.

[8] Gilson, L., Imad, A., Rabet, L., Van Roey, J., Gallant, J.: Real-Time Measurement of Projectile Velocity in a Ballistic Fabric with a High-Frequency Doppler Radar, *Experimental Mechanics*, 2021, Vol.61, pp.533–547, doi:10.1007/s11340-020-00676-4.

[9] Zhao, D. et al.: High-Precision Velocity Measuring System for Projectiles Based on Retroreflective Laser Screen, *Optik*, 2013, Vol.124, No.6, pp.544–548, doi:10.1016/j.ijleo.2012.01.006.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

CFD-BASED NUMERICAL SIMULATION OF THERMAL PROCESSES AND AIRFLOW IN THE PASSENGER COMPARTMENT OF SPECIAL VEHICLE

MILICA IVANOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia mivanovic@mas.bg.ac.rs

SNEŽANA MALIŠIĆ
Technical Test Center, Belgrade, Serbia, snezana.malisic@toc.urio.vs

MARIJA BALTIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia mbaltic@mas.bg.ac.rs

IGOR STAMENKOVIĆ
University of Belgrade, Faculty of Mechanical Engineering, Belgrade, Serbia istamenkovic@mas.bg.ac.rs

Abstract: *Special vehicles frequently encounter numerous challenges to achive and maintain adequate thermal comfort in the passenger compartmen, such as high external temperatures and sun radiation, significant heat generation from onboard equipment and engine, limited space, problems with disposition of supply and exhaust vents, etc. In this paper, a CFD-based numerical simulation of airflow and thermal processes in the passenger compartment of a special vehicle is presented. A three-dimensional model of the enclosed space is developed, which includes internal heat sources from occupants and onboard equipment, as well as external thermal loads introduced via appropriately calculated heat flux boundary conditions. Thermal conditions within enclosed compartments of vehicles for special purpose are governed by complex interactions between airflow, heat transfer and surface radiation. Approach based on CFD can offer detailed insights into airflow dynamics and temperature distribution, which helps address varying thermal comfort levels in enclosed spaces (evaluated using the Predicted Mean Vote (PMV) and Predicted Percentage of Dissatisfied (PPD) indices). Numerical model presented in the paper is validated against experimental data of PPD and PMV values and average cabin temperature. The temperature and velocity distributions obtained through CFD simulation demonstrate the capability of this method to serve as a valuable tool for evaluating thermal comfort, as it enables detailed analysis at any surface, plane or point of interest within the domain. From the other side, CFD introduces the challenge of determining whether a simplified model can simultaneously ensure computational efficiency with reliable predictive capability for the specific HVAC systems in special purpose vehicle, on which some conclusions are given.*

Keywords: *thermal comfort, PMV, PPD, vehicle HVAC system, CFD.*

1. INTRODUCTION

In special-purpose vehicles (such as military, emergency response, security vehicles, etc.) maintaining appropriate thermal conditions presents significantly greater challenges than in conventional passenger cars. These vehicles frequently operate under severe environmental conditions combined with geometric parameters, such as: compact cabin layouts and enhanced structural protection that adds load by means of heat accumulation of the structure (higher thermal mass). Furthermore, high internal heat loads can be generated by electronic equipment along with prolonged mission durations. Occupants frequently wear protective clothing and equipment, which influences heat exchange between the human body and the surrounding environment.

In the commercial automotive sector, the quality of the climate control system is often regarded as an important factor influencing customer satisfaction and vehicle competitiveness. Consequently, considerable effort is devoted to analyzing, optimizing, sizing and validating HVAC systems to achieve high levels of thermal comfort and energy efficiency. The situation is relatively different for special and military vehicles in which operational capability, reliability and crew protection are primary, rather than thermal comfort [1]. As a result, air-conditioning systems have traditionally received less attention during the initial design stages and are frequently incorporated using empirical practices or simplified design procedures. Considering military and urgent systems, already stressful service is enhanced by working in hot environment.

The distribution of temperature, airflow velocity and pressure within an enclosed vehicle compartment is governed by complex fluid flow and heat transfer phenomena. Because of the nature of this interacting mechanisms, predicting the thermal performance of a passenger compartment only through experimental testing

Corresponding author: M. Ivanović;
E-mail: mivanovic@mas.bg.ac.rs
DOI: 10.5937/OTEH26156I

is often time-consuming, expensive and limited to a relatively small number of operating conditions. Under these conditions, conventional methods for HVAC sizing and control are no longer sufficient.

The application of CFD (Computational Fluid Dynamics) analysis in the initial stage of design and afterwards for the evaluation of system work, enables engineers to evaluate multiple design configurations without extensive and expensive experimental testing. Numerical simulations enable investigation of the air temperature influence of supply air diffuser location, airflow, ventilation, radiation and cabin geometry on the resulting thermal and flow field. Additionaly, CFD simulation facilitates the identification of recirculation zones, dead-air regions and temperature non-uniformities that may reduce occupant comfort or impair equipment reliability. Such analyses support engineering decisions during the development and optimization of vehicle HVAC system.

The objective of this paper is to perform a CFD-based numerical simulation of the thermal processes and airflow within the passenger compartment of a special-purpose vehicle, with the aim of validating a simplified geometric and numerical model against experimental data of PMV and PPD value and average cabin temperature. This approach naturally raises the question of the extent to which such a model can be both computationally efficient, cost-effective and accurate.

The investigation focuses on predicting the three-dimensional distributions of air velocity and temperature under representative operating conditions.

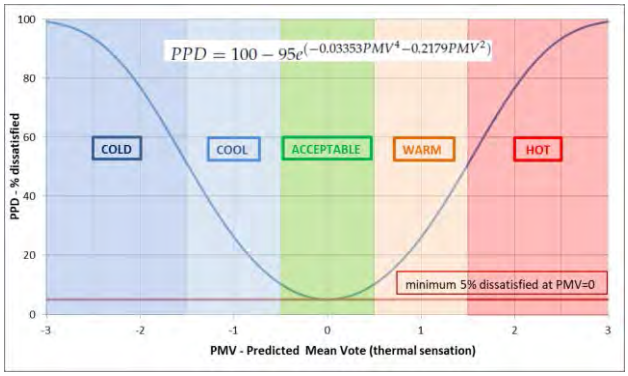
1.1. PPD/PMV indices and standards

Predicted Percentage of Dissatisfied (PPD) is a comfort index that quantifies the expected proportion of occupants dissatisfied with their thermal environment. It is derived directly from the Predicted Mean Vote (PMV) index, which expresses the average thermal sensation of a group on a seven-point thermal sensation scale. This indices are standardized in ISO 7730 standard (Serbian SRPS EN ISO 7730:2008: Ergonomics of the thermal environment - Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. Along with this standard, ANSI/ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy, also rely on PMV/PPD model. ASHRAE 55 defines the acceptable comfort zone as a PMV between -0.5 and +0.5 (meaning ≥ 80% of occupants are satisfied), while ISO 7730 splits this into strict categories (A, B, C) where Class B roughly mirrors ASHRAE's requirements. ISO 7730 also includes local discomfort criteria (radiant asymmetry, draft, floor temperature, thermal stratification).

Air speed has a large effect in warm conditions and is explicitly addressed in provisions of ASHRAE 55 standard. Understanding how these parameters interact helps designers create HVAC systems that achieve occupant satisfaction along with energy efficiency. ASHRAE Standard 55 features a highly adopted Adaptive Comfort Model, which sets dynamic indoor temperature targets based on outdoor climates, while ISO 7730 uses a more rigid approach based on the original Fanger PMV formulas.

The curve from Figure 1 shows PMV/PPD functional dependency with highlight that even under ideal thermal conditions (PMV = 0), at least 5% of people remain dissatisfied which is reflection of individual variation in comfort perception. That is why the PPD index ranges from 5% to 100%, with lower values representing higher comfort levels.

Figure 1. PPD — PMV diagram



The variables of importance to evaluate this indices are the function of design and environmental factors throughout a space:

- temperature (uneven thermal distribution increases discomfort),
- air velocity (low or high velocities can lead to stagnation spots or higher local cooling),
- occupant activity and clothing (metabolic rates (met) and clothing insulation (clo) affect PMV and PPD results),
- humidity (direct impact on evaporative cooling and related effects of perceived temperature),
- asymmetry in radiation (localized discomfort with warm and cold surfaces present).

1.2. Vehicle heat load

Sources of the heat that represent vehicle heat load are ambient and radiation load, engine load, metabolic load and exhaust. Loads can be separated based on the placement into two main groups: outer and internal gains. Outer heat gains are:

- direct Sun radiation,
- diffuse radiation (Sun atmospheric scattering),
- conduction load (ambient),
- ground reflected,
- exhaust heat of the vehicle (rejected by the engine and transferred in the vicinity of the cabin).

Internal heat gains are:

- engine heat (rejected by the engine and transferred into the cabin),
- metabolic heat of the occupants,
- onboard electronic equipment radiation.

3. NUMERICAL MODEL

The researched vehicle has an armored, air-conditioned

cabin and it allows the combat team to operate continuously in a hermetically sealed area protected from NBI (Nuclear-Biological-Chemical) threats. HVAC system completely works in air recirculation mode. Geometry of the simplified cabin model is given in Figure 2.

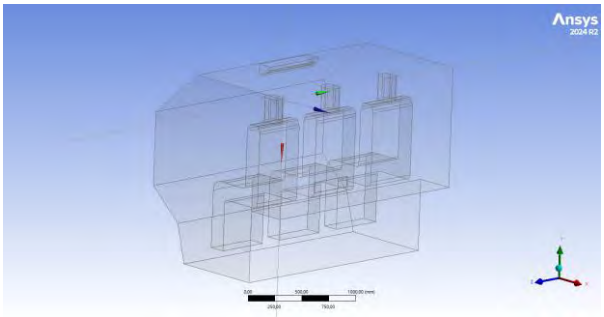


Figure 2. Geometry of the cabin model with occupants

Air circulation in air conditioning mode is carried out through the air distribution system located under the cabin roof and air distribution elements in the front, while exhaust is at the back side of the cabin, which is depicted in Figure 2.

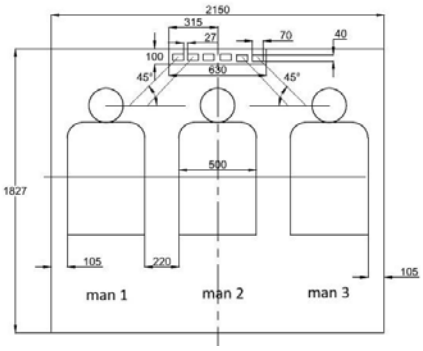


Figure 3. Disposition of supply vents and occupants

Since the simulation of special vehicle that accounts for all potential heat sources is computationally expensive, and moreover, there will be a lot of uncertanty in the final results due to the nature of multiple heat sources, the simplified model has been applied. That model accounts heat gains from the outside (radiation and convection) and electrical equipment radiation, as well as metabolic heat gains, all represented via total heat flux on the particular surfaces (W/m^2).

3.1. Boundary conditions

Experimental condition for air flow is 480 m^3/h, which calculated for the inlet geometry gives v=7,93 m/s. There are six section of the inlet, divided by two for every occupant. The mark of the occupant from the Figure 3 corresponds to the mark of the inlet section. In order to get airflow on every passenger, velocity inlet boundary conditions (BC) are given as components in *x-y-z* direction (Table 2). Turbulence intensity of 5%, and a turbulent-to-laminar viscosity ratio of 10 were set within inlet BC. Inlet temperature is set to 14°C to all sections. Pressure outlet BC was applied at the exit of computational domain. Under this formulation, all flow variables are assumed invariant along the outlet direction, with the exception of the velocity gradient normal to the exhaust plane, while the static pressure at the outlet is set equal to atmospheric pressure.

Table 1. Velocity inlet boundary condition

i	v_x	v_y	v_z
1	$v \cdot \sin(45^\circ) \cdot \cos(45^\circ)$	$-v \cdot \cos(45^\circ)$	$v \cdot \sin(45^\circ) \cdot \sin(45^\circ)$
2	$v \cdot \sin(45^\circ)$	$-v \cdot \cos(45^\circ)$	0
3	$v \cdot \sin(45^\circ) \cdot \cos(45^\circ)$	$-v \cdot \cos(45^\circ)$	$-v \cdot \sin(45^\circ) \cdot \sin(45^\circ)$

Heat flux boundary conditions applied to the vehicle cabin surfaces (except windshield) were determined using the sol-air temperature method (simplifying heat gain calculations by combining ambient air temperature, solar radiation and thermal radiation into a single equivalent outdoor temperature) for all surfaces except windshield. This condition is combined with one-dimensional steady-state conduction through the composite armored and insulatied wall. The daily solar insolation for is Belgrade in July is 6.4 kWh/m²/day (Global horizontal) [2], which converted to peak current flux gives 657 W/m². Considering more conservative approach of "clear-sky" peak values, which other autors used (930 W/m² in [3], 875 W/m² in [4], and 1120 W/m² in [5]), we decided to use I_{solar} = 900 W/m². The absorbed solar flux was calculated using a paint solar absorptivity of α = 0.85, derived from the absorbed-to-incident irradiance ratio reported in [5], which gives absorbed flux on the outer surface of 765 W/m². The external sol-air temperature ($T_{sol\text{-}air}$) is calculated using the ASHRAE Handbook heat balance relationship:

$$T_{sol-air} = T_o + \frac{\alpha \cdot I_t}{h_o} - \frac{\varepsilon \cdot \Delta R}{h_o} \tag{1}$$

With a safety margin for summer period the third member is omitted. The external convective heat transfer coefficient was calculated using the correlation from [1], yielding $h_{ext} \approx 29.5$ W/m²K at a vehicle speed of 30 km/h. The sol-air temperature was then used together with the thermal resistance of the 30 mm armor (λ = 15 W/m·K) [6] and 16 mm insulation (λ = 0.035 W/m·K) [7], to determine the interior wall heat flux driven by the difference between the sol-air temperature and the target cabin interior temperature of 26 °C. To account for reduced solar exposure, the effective irradiance on the sides and rear walls was assumed to be around 70% of the roof value.

The total incident solar heat flux on the tilted windshield (at a tilt angle of 40°) was determined by considering direct, diffuse, and reflected radiation components under peak summer conditions in Belgrade at 13:00 h. The direct radiation component was computed by modulating the Direct Normal Irradiation (DNI) with the cosine of the solar angle of incidence (θ), while the Diffuse Horizontal Irradiation (DHI) and Global Horizontal Irradiation (GHI) were scaled using the respective sky and ground view factors, incorporating a standard asphalt albedo of ρ_g = 0.15. To evaluate the actual thermal energy penetrating the cabin through the athermic glass, the calculated total

incident flux was coupled with a standard Solar Heat Gain Coefficient (SHGC=0.40).

Radiative heat transfer from onboard electronic equipment was estimated using the Stefan–Boltzmann relation with an emissivity of 0.9, surrounding temperature of 26 °C, and equipment surface temperatures of 50°C. Heat fluxes generated by the occupants are accounted according to the chosen one for estimating PMV/PPD values during experimental testing. Resulting heat fluxes and surface emissivities implemented as BC are given in the Table 2.

Table 2. Surface heat fluxes and emissivities

Section	Heat flux (W/m^2)	Emissivity
roof	90	0,9
windshield	180	0,9
rear	65	0,9
bottom	30	0,85
front (el.equipment)	150	0,9
sides	65	0,9
man 1	70	0,95
man 2	70	0,95
man 3	70	0,95

3.2. Mesh convergence

Discretization, flow and thermal calculations were performed using ANSYS Fluent [8][8]. A polyhedral mesh was formed with local sizing/refinement in the zones of the ventilation openings and on the occupants. Along the walls of the cabin and around bodies, 4 layers of the boundary layer with a growth factor of 1.2 were formed. The independence of the solution from the mesh density was confirmed with the analysis of three levels of mesh refinement (0.8, 1.2 and 2.2 million cells), where the difference in average cabin temperature was less than 0,3 % between medium and fine mesh (volume averaged temperatures in fluid domain were: 297,84 K, 299,31 K and 300,01K), so medium mesh was adopted for further simulations.

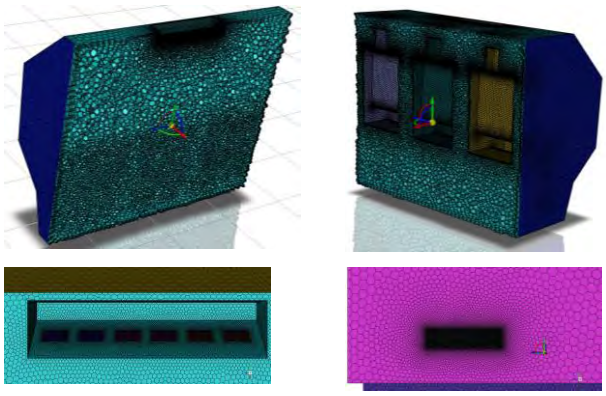


Figure 4. Details of the mesh

3.3. Methods

Radiative heat transfer from the heat sources was modeled using the Discrete Ordinates (DO) radiation model, whereas buoyancy-induced airflow arising only from density gradients within the cabin was accounted with Full buoyancy effects.

Turbulence closure was achieved through a modified two-equation Realizable k-ε model, along with Enhanced Wall Treatment and with Thermal Wall Function.

In the initial phase (first 300 iterations), the pressure-velocity coupling was managed via the SIMPLEC algorithm, combined with First-Order Upwind schemes for the momentum, energy and turbulence equations to establish a stable, non-oscillatory initial flow field. Pressure discretization was done with Body-force-weighted scheme.

For the secondary phase, the solution configuration was advanced to the Coupled algorithm to simultaneously solve the mass and momentum equations. In order to more accurately capture the buoyancy-induced flows driven by significant temperature gradients within the cabin, the pressure discretization was upgraded to the PRESTO! (PREssure STaggering Option) scheme, which is suited for natural convection with higher Rayleigh number. Spatial discretization for the remaining transport variables was elevated to the Second order upwind scheme.

Given the numerical instability inherent in coupled forced and natural convection regimes, the pseudo-transient formulation was included. The Time scale factor was systematically managed, initiating at lower bounds (1e-3) to damp transient numerical fluctuations during the transition to higher order discretization. The residuals stably converged below the thresholds of 1e-3 for continuity and 1e-6 for energy.

4. RESULTS

4.1. Experimental results

Experimenntal testing was conducted after system reached a steady-state thermal condition. Air temperature and velocity, relative humidity (RH), cabin pressure and radiant temperature were measured using sensors for each quantity, over a duration of 5 minutes with a sampling frequency of 1 s. The sensors were positioned at multiple points on different cabin segments (the height at which the sensors are placed is 1,1 m and 0,1 m from the floor of the vehicle as defined in the ISO 7730). This measurement protocol provided a sufficient number of time points (300 samples per sensor).

During the measurement, it was observed that the speed of air flow in the zone around the head is uneven. Unfortunately, technical possibilities and space dictate limitations. Those limitation cannot be changed.

Based on the measured values and using assumed values of metabolic rate, clothing thermal insulation (normal business wear; 0,154 m²K/W clo=0,57) and activity (sitting down, light activity; 70 W/m²; met=1,2), PMV and PPD indices were calculated according to Fanger's thermal comfort model, as defined in the ISO 7730 standard. The air speed was measured in front and was 2 m/s in the narrow zone where it was measured.

The results of average values, as well as extreme values are given in Table 3.

Table 3. Experimental data of temperatures and RH

	t_1, °C	t_2, °C	t_3, °C	%, RH
Average	26,8	25,8	27	36,1
Max.	27	26,1	27,1	39,4
Min.	26,5	25,2	26,6	35,2

Measured values resulted in values of PMV and PPD presented in Table 3.

Table 3. Obtained values of PMV and PPD

	PMV	PPD
Average	0,8	18,5
Maximum	1	22,1
Minimum	0,7	15,3

4.2. CFD results

For the purpose of simplification of the numerical model, relative humidity (RH) was not included in the CFD analysis. Experimental data of this variable showed only 4,2% difference between minimal and maximum value. For the calculation of PMV/PPD values, the average RH was used. Obtained value of average volume in fluid domain are 26,2 °C, while the temperatures at the surface of the passengers are given in Table 4. Surface averaged velocity for the plane created in front of the bodies (at the distance of 10 cm from the chest area) is 0,433 m/s.

Table 4. Numerical results for the body surface of the occupants

	Man 1	Man 2	Man 3
Surface averaged temperature, °C	33,32	33,19	33,44
Radiation temperature, °C	34,75	34,69	34,71
Total heat transfer rate, W	175,02	166,62	169,85

For the calculation of the PMV and PPV value the CBE Thermal Comfort Tool is used. That is web-based calculator developed by the Center for the Built Environment at UC Berkeley and is in compliance with a ASHRAE 55, ISO 7730 and EN 16798-1 standards.

Obtained PMV value is 0,83, while PPD is 20% according to EN 16798-1 standard.

The average temperatures on the passenger's body surface, obtained in the simulation, are in good agreement with the range considered physiologically neutral and comfortable for the human body in the literature (30-34°C) [9]. Given that the obtained average body surface temperatures in this simulation are within this recommended range, it can be concluded that the simulated ventilation conditions ensure a satisfactory thermal state of the passenger's body on average, although, as is shown at the Figure 5, local variations (especially in the head and lower extremities) may deviate from this neutral value.

The temperature distribution across the surfaces of the three occupants results reveal a clear vertical stratification of body-surface temperature, with the highest values at the head. This localized heating is consistent with the elevated temperatures identified on the interior roof surface, indicating that radiative heat transfer from the hot roof panel directly contributes to the thermal loading of the occupants head. In contrary, the back and upper torso region exhibits noticeably lower temperatures, in the range of approximately 29–35°C, which can be attributed to the direct impingement of the cold air jet discharged from the vents, as it can also be observed in the velocity contours from Figure 5. Moving further down the body, the lower torso and thigh region shows intermediate and elevated temperaturee, suggesting reduced local ventilation efficiency in this area and a corresponding accumulation of heat. Overall, this non-uniform thermal pattern across the body demonstrates that occupant thermal comfort cannot be adequately captured by a single average body temperature and supports the use of segment based approach in order to to properly characterize localized thermal sensation.

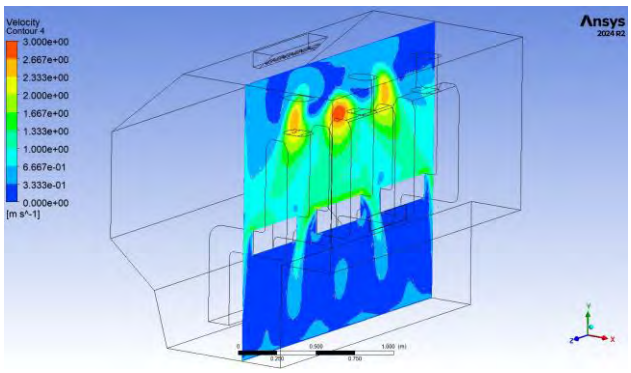


Figure 5. Velocity contours in the plane in front of the occupants

The temperature distribution on the interior surfaces of the cabin is shown in Figure 6. The results show pronounced spatial unevenness of the temperature field, with the highest values, up to 59.1°C, recorded on the windshield next to the cabin roof. These indicates that the windshield is the dominant source of radiant heat load inside the cabin, which is in line with physical expectations for a vehicle exposed to direct solar radiation. That suggests thermal insulation at the junction between the roof and the windshield, where thermal bridges are likely to be formed.

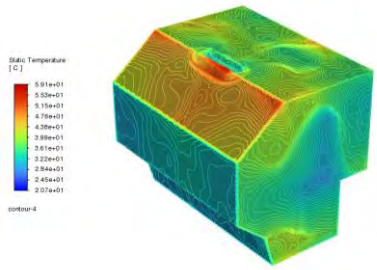


Figure 6. Temperature distribution on cabin interior surfaces (walls)

Considering the given conditions of speed angles, passengers can feel discomfort expressed in the form of uneven temperature on different parts of the body, primarily in the area of the legs and crown of the head. The applied heat flux has been increased on the safety side, so that in reality lower values are expected, and therefore less pronounced discomfort.

The distribution of air flow velocity in the longitudinal section of the cabin is shown by the vector field in Figure 7. The jet is directed downwards towards the central part of the cabin. After entering, the jet maintains a relatively high speed along a narrow vertical channel, after which it expands and gradually slows down due to mixing with the surrounding, much calmer air in the lower part of the cabin, where low flow speeds (below 2 m/s) dominate. This distribution is a direct consequence of the one-sided position of the inlet because entire airflow is supplied from one, upper side of the cabin, there is no symmetrical or multi-sided flow distribution along the longitudinal axis of the cabin, which leads to a pronounced velocity gradient between the central zone directly below the inlet and the front and rear of the cabin, where zones of lower velocity and potential stagnation are formed.

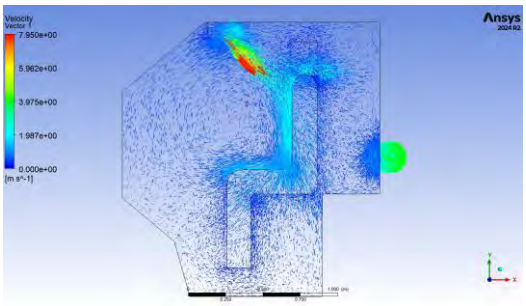


Figure 7. Velocity vector in cabin mid cross section

5. CONCLUSION

In this paper, a CFD analysis of thermal comfort and air flow in the cabin of a vehicle with three cabin crew members was carried out. The simulation included a cabin with a unique arrangement of air supply exclusively from the upper side of the cabin, which enabled a detailed analysis of the impact of spatially asymmetric configuration on the distribution of speed, temperature and thermal comfort indices of passengers.

The thermal comfort indices PMV and PPD calculated for the observed case were 0,83 and 20%, respectively, which is in good agreement with data obtained through experimental testing.

The results of this research indicate that the assessment of thermal comfort based solely on global, spatially averaged indices (PMV, PPD, average temperature) and simplified geometry does not fully include local variations of the thermal and velocity field inside the cabin, especially in the zones of direct interaction with air jets (head, chest) and solar heating zones (windshield and roof).

The obtained results contribute to a better understanding of airflow organization and heat transfer processes and provide a basis for future optimization of HVAC system performance, energy efficiency and occupant thermal comfort in special purpose vehicles.

References

[1] Brooks, M.-A. (2016). *Thermal modeling for light armored vehicle* [Master's thesis, University of Western Ontario]. https://ir.lib.uwo.ca/etd/3745

[2] BAESS labs solar irradiance calculator: https://www.baess.app/tools/solar-irradiance-calculator

[3] Fujita, A., Kanemaru, J., Nakagawa, H., & Ozeki, Y. (2001). Numerical simulation method to predict the thermal environment inside a car cabin. *JSAE Review*, 22(1), 39–47.

[4] Park et al. (2026). Comprehensive parametric study of cabin thermal comfort [proveri pun naslov]. *Applied Sciences*, 16(8), 3964. https://doi.org/10.3390/app16083964

[5] Pryor, R., Ditty, D., Mao, W., Rynes, P., & Smith, C. (2009). Development of a validated thermal model of AC performance in a protected combat ground vehicle. *Proceedings of GVSETS 2009*, NDIA/SAE, 1–7.

[6] Elasto Proxy. Military Vehicle Thermal Insulation. Technical reference (manufacturer data). https://www.elastoproxy.com/military-vehicle-thermal-insulation/

[7] Engineering ToolBox. Mineral Wool Insulation — Thermal Conductivity. Technical reference. https://www.engineeringtoolbox.com/mineral-wool-insulation-k-values-d_815.html

[8] https://www.ansys.com/

[9] Zhang, L., et al., Numerical Analysis of the Relationship Between Human Skin Temperature and Thermal Comfort. Proceedings of the ACM (2022). DOI: 10.1145/3501409.3501619

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

ACCURACY ASSESSMENT OF GROUND TARGET COORDINATES DETERMINED BY DIFFERENT ACQUISITION SYSTEMS

SRĐAN STOJAKOVIĆ
Military Geographical Institute – „General Stevan Bošković", Belgrade, sstojakovic81@gmail.com

SINIŠA DROBNJAK
Military Geographical Institute – „General Stevan Bošković", Belgrade, sinisa.drobnjak@vs.rs

NOVAK ROGANOVIĆ
Military Geographical Institute – „General Stevan Bošković", Belgrade, novakroganovic1999@gmail.com

SAŠA BAKRAČ
Military Geographical Institute – „General Stevan Bošković", Belgrade, sbakrac2017@gmail.com

Abstract: *Accurate determination of ground target coordinates is an important requirement in military, security and geospatial applications. Different target acquisition systems, such as electro-optical sensors, laser rangefinders, GNSS/INS units, total stations and photogrammetric platforms, can provide target coordinates, but their results are not directly comparable without a clearly defined testing methodology. This paper presents methods for assessing the accuracy of ground target coordinates determined by different acquisition systems. The proposed approach is based on the comparison of measured target coordinates with independently determined reference coordinates on a dedicated geodetic test field. In addition to direct coordinate comparison, repeated observations, relative geometric checks and operational scenario testing are considered. The paper also proposes statistical indicators for reporting results, including coordinate residuals, horizontal and three-dimensional position errors, RMSE values, systematic offsets and empirical percentiles. The presented methodology provides a basis for objective evaluation, verification and comparison of target acquisition systems under controlled and operational conditions.*

Keywords: *target acquisition; ground targets; coordinate accuracy; RMSE; test field.*

1. INTRODUCTION

Accurate determination of ground target coordinates is an important requirement in military, security, geospatial and reconnaissance applications. Target coordinates are used for surveillance, fire support, command and control, navigation and integration of sensor data into geospatial information systems. They can be determined by different acquisition systems, including electro-optical and infrared sensors, laser rangefinders, GNSS/INS units, digital compasses, total stations, photogrammetric systems, unmanned aerial platforms and integrated multi-sensor systems.

Although these systems may provide the same type of output, their accuracy may differ significantly. The final coordinate result depends not only on the nominal accuracy of individual sensors, but also on the complete measurement process. Important error sources include sensor position, orientation accuracy, distance and angular measurements, target identification, atmospheric conditions, coordinate transformation, time synchronization and operator influence.

The terms accuracy and precision are often used interchangeably, although they have different meanings. According to the ISO 5725 series, accuracy includes both trueness and precision, while precision refers to the closeness of repeated measurement results under specified conditions [1]. This distinction is important because a target acquisition system may produce highly repeatable coordinates that are still affected by systematic errors. In the military context, this problem is closely related to Target Location Error (TLE), which represents the difference between the actual target location and the location determined by the acquisition process [2].

A reliable assessment of target coordinate accuracy requires an independent reference framework. This can be achieved by establishing a geodetic test field with clearly defined target points whose coordinates are determined by higher-accuracy geodetic methods. The coordinates obtained by the tested acquisition systems can then be compared with reference coordinates. Similar principles are used in positional accuracy assessment of geospatial data, where independent checkpoints and RMSE indicators are commonly applied [3]. The interpretation of results should also consider measurement uncertainty, following general principles given in the Guide to the Expression of Uncertainty in Measurement [4].

The aim of this paper is to present a concise methodological framework for testing the accuracy of ground target coordinates determined by different acquisition systems. The paper discusses the concept of a

Corresponding author: S. Stojaković;
E-mail: sstojakovic81@gmail.com
DOI: 10.5937/OTEH26157S

reference test field, main testing methods and statistical indicators required for objective evaluation, verification and comparison of acquisition systems under controlled and operational conditions.

2. TARGET ACQUISITION SYSTEMS AND ERROR SOURCES

Ground target coordinates can be determined using different types of acquisition systems. In general, these systems combine information about the position and orientation of the sensor with measurements directed towards the target. Depending on the system concept, target coordinates may be obtained from optical observation, laser distance measurement, angular measurements, GNSS/INS navigation data, photogrammetric processing or a combination of several sensors. Typical acquisition systems include handheld or tripod-mounted electro-optical devices with laser rangefinders, vehicle-mounted observation systems, airborne or UAV-based EO/IR platforms, GNSS/INS-based multi-sensor systems, total stations and photogrammetric systems. Some of the mentioned acquisition systems are shown in Figure 1. Although these systems differ in design and operational use, the final result is usually expressed as target coordinates in a selected coordinate reference system.

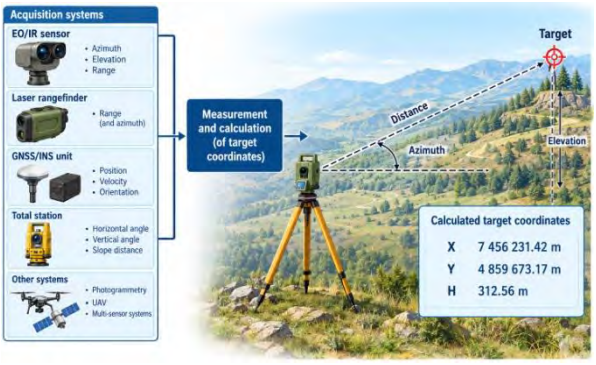


Figure 1. Target coordinate determination system

Studies dealing with target geo-location and acquisition systems show that the final coordinate accuracy is affected by the complete measurement chain, not only by individual sensor specifications. In airborne optoelectronic platforms, important error sources include sensor measurement errors, installation errors and transformations between different coordinate frames [5]. For systems based on laser rangefinders and inertial or attitude sensors, orientation error is often one of the dominant factors influencing the final target position [6]. Laser rangefinder-based airborne EO systems also show that platform position and distance measurement errors have a significant effect on target geo-location accuracy [7]. In image-based UAV systems, geolocation errors become more evident at longer distances and are strongly influenced by GNSS/INS accuracy, camera parameters and target elevation information [8].

The main error sources can be grouped into four categories. Sources of error in target coordinate detrmination are shown in Figure 2. The first category is related to the position of the acquisition system. Any error in sensor or platform coordinates is transferred directly to the computed target coordinates. This is especially important for mobile and airborne systems, where the sensor position is usually determined by GNSS or GNSS/INS integration.

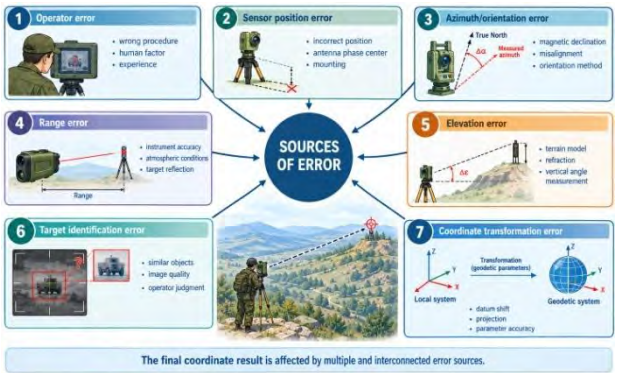


Figure 2. Sources of error in target coordinate detrmination

The second category includes orientation-related errors. Errors in azimuth, pitch, roll or sensor alignment may cause significant horizontal and vertical displacement of the target coordinates. Even a small angular error can produce a large transverse error at longer distances, which makes orientation accuracy one of the key factors in target coordinate determination.

The third category includes observation errors related to the target itself. These errors may originate from laser distance measurement, horizontal and vertical angle measurement, target identification, image resolution, camera calibration, atmospheric conditions, visibility and operator aiming. In image-based systems, additional errors may be caused by feature extraction, image matching and georeferencing.

The fourth category includes computational and reference system errors, such as coordinate transformation errors, incorrect geodetic datum parameters, height system inconsistencies and time synchronization errors. These errors are often difficult to detect during field use, but they may introduce systematic offsets in the final target coordinates.

For accuracy testing, it is therefore necessary to evaluate the complete acquisition process. The main error sources that influence ground target coordinate determination are summarized in Table 1.

Table 1: Main error sources in ground target coordinate determination

Category	Error source	Typical influence on target coordinates
Category 1 - position of the acquisition system	Sensor or platform position	Direct translation of the computed target coordinates
Category 2 - orientation-	Azimuth and orientation errors	Horizontal displacement

related errors		increasing with target distance
	Vertical angle and pitch errors	Height error and range-dependent position error
Category 3 - errors related to the target	Distance measurement error	Error mainly along the line of sight
	Target identification and aiming	Random or systematic targeting error
	Camera calibration and image processing	Image-based coordinate error
Category 4 - computational and reference system errors	Atmospheric and visibility conditions	Reduced observation quality and range accuracy
	Coordinate transformation error	Systematic coordinate offset
	Time synchronization error	Position error for moving platforms or targets
	Operator influence	Variations in aiming and measurement procedure

3. REFERENCE TEST FIELD CONCEPT

A reliable assessment of ground target coordinate accuracy requires a reference framework that is independent of the tested acquisition system. For this purpose, a dedicated geodetic test field can be established, consisting of clearly defined and stable target points with accurately determined reference coordinates. These coordinates should be obtained using higher-accuracy geodetic methods, such as static GNSS observations, total station measurements, precise levelling or a combination of independent techniques.

The concept of such a test field should be closer to a reconnaissance and target acquisition polygon than to a conventional firing range. Its primary purpose is not to test weapon effects, but to evaluate how accurately different systems can detect, identify and determine the coordinates of ground targets. Therefore, the field should combine geodetic accuracy requirements with operationally realistic observation conditions.

Publicly available examples relevant to this concept can be found in remote sensing calibration and validation sites, UAV target geolocation experiments and military sensor trials. The Baotou Comprehensive Calibration and Validation Site in China is equipped with artificial standard targets, including portable and permanent targets, and supports long-term calibration and validation of remote sensing sensors [9]. A UAV long-range oblique photography experiment presented by Liu et al. used three ground targets at slant ranges of approximately 48 km, 52 km and 49 km, while the ground-truth target coordinates were measured by a geodetic GNSS receiver with an error smaller than 0.1 m [8]. In Europe, the FGI Sjökulla Aerial Test Range in Finland includes permanent ground control points with precisely known coordinates, test-bar targets and reflectance panels for geometric and radiometric calibration of imaging systems [10]. Military sensor trials, such as NATO SET-250, also show the importance of observing military targets from different angles and configurations [11].

Based on these examples, a military-oriented reference test field should include open terrain for long-range line-of-sight observations, semi-urban or urban structures for complex backgrounds and partial occlusions, and areas suitable for tripod-mounted, vehicle-mounted and UAV-based acquisition systems. Targets should be distributed at different distances, azimuths and height differences relative to possible observation stations. This geometry makes it possible to evaluate how coordinate accuracy changes with range, viewing direction and vertical angle.

The test field should include several types of targets. Geodetic targets, such as marked points, prisms, retroreflectors and high-contrast panels, are suitable for reference coordinate determination and for testing visible-spectrum and laser-based systems. Photogrammetric targets, such as coded or cross-shaped ground control markers, are useful for UAV and image-based systems. For thermal imaging systems, thermally contrasted panels, heated plates, vehicle-size thermal targets and standard four-bar targets may be used. Four-bar targets are commonly used in thermal imager testing, especially for MRTD-based performance assessment [13]. Target acquisition performance for electro-optical and thermal systems can also be evaluated by field trials and models that consider detection, recognition and identification criteria [14], while thermal imager range performance depends on target characteristics, atmosphere, background, sensor parameters and selected perception criteria [15]. Proposed geodetic test field for accuracy assessment of target coordinates is shown on Figure 3.

Figure 3. Geodetic test field for accuracy assessment of target coordinates

In a reconnaissance-oriented test field, the target set should also include operationally representative objects, such as vehicle silhouettes, mock vehicles, shelters, observation posts, building corners, road intersections and personnel-size targets. These objects should not be treated only as visual targets, but as coordinate targets with precisely defined reference points, for example the centre

of a panel, the centre of a thermal target, a marked point on a vehicle silhouette or a selected structural corner. This is important because ambiguity in target definition may become a significant part of the final coordinate error.

A possible layout may include several observation stations with known coordinates and targets grouped into short, medium and long-range zones. Short-range targets, for example from 100 m to 500 m, are suitable for testing aiming, angular measurement and close-range coordinate determination. Medium-range targets, from about 1 km to 3 km, are suitable for handheld and vehicle-mounted EO/IR systems with laser rangefinders. Long-range targets, beyond 5 km, are useful for analysing the influence of orientation, ranging, atmospheric conditions and UAV-based observation geometry.

The main purpose of the reference test field is to provide targets whose coordinates are known with an uncertainty significantly smaller than the expected errors of the tested acquisition systems. The coordinates obtained by the tested systems can then be compared with reference coordinates. The same field can also be used for repeated observations, relative geometric checks and operational scenario testing. This type of configuration links geodetic traceability with operationally realistic target acquisition conditions.

4. TESTING METHODS

Several complementary methods can be used for testing the accuracy of ground target coordinates determined by different acquisition systems. The selection of methods depends on the type of system, observation distance, target type, required accuracy and operational conditions. Testing should include both comparison with reference coordinates and analysis of repeatability under defined conditions. A review of target geo-location algorithms for aerial remote sensing systems shows that the final target position depends on the sensor position and attitude, imaging geometry, target elevation, coordinate transformations and the applied geo-location model [15].

4.1. Direct comparison with reference coordinates

Direct comparison with reference coordinates is the basic method for accuracy assessment. The tested acquisition system determines the coordinates of a selected target, and the obtained coordinates are compared with independently determined reference coordinates.

This method directly estimates the coordinate error of the tested system and enables the detection of both random and systematic effects. It is conceptually similar to positional accuracy assessment of geospatial data, where independent checkpoints and coordinate residuals are used for accuracy evaluation [3]. The main requirement is that the uncertainty of the reference coordinates must be significantly smaller than the expected errors of the tested acquisition system. Comparison of measured and reference coordinates is shown in Figure 4.

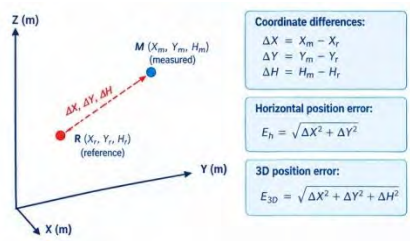


Figure 4. Comparation of measured and reference coordinates

4.2. Repeated observations

Repeated observations of the same target are used to evaluate the repeatability of the acquisition system. The same target is observed several times under identical or similar conditions, and the dispersion of the obtained coordinates is analysed. This method is useful for determining short-term stability, operator influence, aiming repeatability and measurement noise.

However, repeated observations cannot by themselves prove coordinate accuracy. A system may provide highly repeatable results that are still affected by a systematic error. Therefore, repeatability testing should always be combined with comparison against reference coordinates. This distinction is consistent with the ISO approach, where precision describes the closeness of repeated measurements, while accuracy includes both trueness and precision [1].

Repeated observations should be performed for different target distances and target types. Short-range observations may be used to evaluate aiming and angular measurement effects, while longer-range observations are more sensitive to orientation, distance measurement, target elevation and atmospheric influences. Long-range oblique reconnaissance studies also show that target geo-location accuracy is strongly affected by observation geometry and by the availability or uncertainty of target elevation information [16].

4.3. Relative geometric checks

Relative geometric checks are based on the comparison of geometric relationships between targets. Instead of analysing only absolute coordinates, distances, directions, angles or height differences between pairs of targets are computed from the measured coordinates and compared with the corresponding reference values.

This method is useful for detecting scale effects, angular effects, local geometric distortions and range-dependent errors. It can also help identify whether the dominant error component is related to distance measurement, orientation or coordinate transformation. Relative checks are especially useful when several targets are observed from the same station or when the same target group is observed by different acquisition systems.

The limitation of this method is that it may not detect a common translation of all target coordinates. Therefore, relative geometric checks should be used as a complementary method, not as a replacement for direct

comparison with reference coordinates.

4.4. Operational scenario testing

Operational scenario testing evaluates the performance of acquisition systems under conditions similar to real use. The test should include different observation distances, azimuths, vertical angles, target types, visibility conditions, operators and observation platforms. Depending on the system, testing may be performed from tripod-mounted stations, vehicles, UAVs or other platforms.

The purpose of this method is to determine how the accuracy changes when observation conditions are no longer ideal. For EO/IR systems with laser rangefinders, tests should include different visibility and range. For thermal systems, target acquisition performance is commonly related to detection, recognition and identification criteria, while thermal imager performance depends on target characteristics, atmosphere, background and sensor parameters [12], [13], [14]. For GNSS/INS-based systems, tests should include open terrain, semi-urban areas and areas with reduced GNSS quality. UAV-based image acquisition systems should be tested with different flight geometries, image overlaps, camera orientations and target distances. Multi-view and long-range UAV target geo-location experiments show that observation geometry, sensor accuracy and reference target coordinates have a significant influence on the final target position [8], [16].

Operational testing should also include repeated observations by different operators. This makes it possible to estimate the influence of target selection, aiming procedure and interpretation of the target reference point. The results should be reported separately for different scenarios, because one overall accuracy value may hide important differences between short-range, medium-range, long-range, optical, thermal and image-based observations.

4.5. Combined testing approach

No single method is sufficient for a complete assessment of target coordinate accuracy. Direct comparison with reference coordinates should be the primary method, because it provides the most direct estimate of coordinate error. Repeated observations provide information about precision and stability. Relative geometric checks help identify geometric and systematic effects, while operational scenario testing shows how the system performs under realistic field conditions.

For a reconnaissance-oriented reference test field, the recommended testing procedure is therefore based on a combination of these methods. Each tested system should observe the same set of reference targets under predefined scenarios. Coordinate differences, horizontal and three-dimensional errors, RMSE values, systematic offsets and empirical percentiles should then be calculated and reported. The interpretation of results should also consider the uncertainty of reference coordinates and the general principles of measurement uncertainty evaluation [4].

5. STATISTICAL INDICATORS AND REPORTING

The results of target coordinate accuracy testing should be reported using a limited set of clear and comparable statistical indicators. The basic quantities are coordinate residuals, calculated as differences between the coordinates determined by the tested acquisition system and the corresponding reference coordinates. These residuals are used for the analysis of systematic offsets, random dispersion and overall position error.

Mean coordinate residuals should also be calculated, because they indicate possible systematic effects in the coordinate solution. Standard deviations of repeated observations provide information about precision and short-term stability. However, standard deviation alone is not sufficient for accuracy assessment, because it does not describe the offset from reference coordinates.

For operational evaluation, empirical percentiles of horizontal or three-dimensional position errors are also useful. The values $(P_{\{50\}})$, $(P_{\{90\}})$and $(P_{\{95\}})$ indicate the error level below which 50%, 90% and 95% of the observed errors are located. These indicators are practical for comparing different systems and scenarios, especially when the error distribution is not perfectly normal.

The results should not be reported only as one overall value. They should be grouped by acquisition system, target distance, target type, observation platform, operator and test scenario. This makes it possible to identify whether the coordinate error is mainly related to distance, orientation, target identification, sensor type or operational conditions. The uncertainty of reference coordinates should also be stated, because it directly affects the interpretation of the obtained differences [4].

6. PROPOSED TESTING PROCEDURE

he proposed testing procedure is based on the use of a reconnaissance-oriented reference test field and a predefined set of observation scenarios. The procedure should be applicable to different acquisition systems, including tripod-mounted, vehicle-mounted, UAV-based, electro-optical, thermal, laser rangefinder, GNSS/INS and photogrammetric systems [18].

In the first step, reference target points and observation stations should be established. Each target must have a clearly defined reference point, such as the centre of a panel, the centre of a thermal target, a marked point on a vehicle silhouette or a selected structural corner [19]. The coordinates of all targets and observation stations should be determined by higher-accuracy geodetic methods. The coordinate reference system, height system, survey method and estimated uncertainty should be documented [20].

In the second step, test scenarios should be defined. Targets should be grouped according to distance, target type and observation conditions. A basic configuration may include short-range targets from 100 m to 500 m, medium-range targets from 1 km to 3 km and long-range targets beyond 5 km. The test should include optical

targets, geodetic targets, photogrammetric markers, thermally contrasted targets and operationally representative objects such as mock vehicles, shelters, observation posts and building corners.

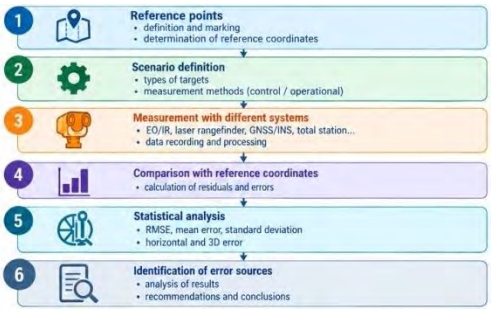


Figure 5. Flowchart of the proposed testing procedure

Flowchart of the proposed testing procedure is shown on Figure 5. In the third step, each acquisition system should determine the coordinates of the selected targets according to the same predefined procedure. For each system, the observation position, sensor configuration, operator, time of observation, atmospheric conditions, visibility, target type and measurement mode should be recorded. Repeated observations should be performed in order to evaluate precision and short-term stability [21].

In the fourth step, the obtained coordinates should be compared with the reference coordinates. Coordinate residuals, horizontal and three-dimensional errors, RMSE values, mean residuals, standard deviations and empirical percentiles should be calculated [22]. The results should be reported separately for different systems, target distances, target types, operators and test scenarios [18].

In the final step, the results should be analysed in order to identify dominant error sources. Direct comparison with reference coordinates provides the main estimate of accuracy, repeated observations describe precision and stability, relative geometric checks indicate possible scale or orientation effects, and operational scenario testing shows the behaviour of the system under realistic field conditions. The interpretation of the results should also include the uncertainty of reference coordinates and the limitations of the applied test configuration [4].

7. CONCLUSION

The accuracy of ground target coordinates determined by different acquisition systems cannot be reliably assessed only from manufacturer specifications or repeated observations. A complete evaluation requires comparison with independently determined reference coordinates, supported by repeatability analysis, relative geometric checks and testing under operationally realistic conditions.

The proposed methodology is based on a reconnaissance-oriented geodetic test field with clearly defined targets, known reference coordinates and different observation geometries. Such a field enables the testing of tripod-mounted, vehicle-mounted, UAV-based, electro-optical, thermal, laser rangefinder, GNSS/INS and photogrammetric acquisition systems under comparable conditions.

Direct comparison with reference coordinates should be the primary method for estimating coordinate accuracy, while repeated observations provide information about precision and stability. Relative geometric checks can help detect scale, orientation and transformation effects, and operational scenario testing allows the assessment of system performance under real field conditions.

The results should be reported using coordinate residuals, horizontal and three-dimensional errors, RMSE values, systematic offsets, standard deviations and empirical percentiles. This provides a consistent basis for objective evaluation, verification and comparison of different target acquisition systems. The proposed approach links geodetic traceability with operational requirements and may support the development of dedicated military geodetic test fields for reconnaissance and target coordinate accuracy assessment.

References

[1] ISO 5725-1:2023, Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions, International Organization for Standardization, Geneva, 2023.

[2] U.S. Air Force, Air Force Pamphlet 14-210, Intelligence Targeting Guide, Chapter 13: The Target Location, 1998.

[3] ASPRS, Positional Accuracy Standards for Digital Geospatial Data, Edition 2, American Society for Photogrammetry and Remote Sensing, 2023.

[4] JCGM 100:2008, Evaluation of measurement data — Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, 2008.

[5] SUN, H., JIA, H., WANG, L., XU, F., LIU, J.: Systematic Error Correction for Geo-Location of Airborne Optoelectronic Platforms, Applied Sciences, 11(22), 11067, 2021.

[6] ZHAO, L., GUAN, D., LANDRY, R.JR., CHENG, J., SYDORENKO, K.: An Accurate and Fault-Tolerant Target Positioning System for Buildings Using Laser Rangefinders and Low-Cost MEMS-Based MARG Sensors, *Sensors*, 15(10), 27060–27086, 2015.

[7] ZHANG, H., QIAO, C., KUANG, H.: Target Geo-Location Based on Laser Range Finder for Airborne Electro-Optical Imaging Systems, *Optics and Precision Engineering*, 27, 8–16, 2019.

[8] LIU, C., DING, Y., ZHANG, H., XIU, J., KUANG, H.: Improving Target Geolocation Accuracy with Multi-View Aerial Images in Long-Range Oblique Photography, *Drones*, 8(5), 177, 2024.

[9] LI, C.R., TANG, L.L., MA, L.L., ZHOU, Y.S., GAO, C.X., WANG, N., LI, X.H., WANG, X.H., ZHU, X.H.: Comprehensive Calibration and Validation Site for Information Remote Sensing, The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, XL-7/W3 (2015) 1233–1240.

[10] EROS Cal/Val Center of Excellence, FGI Sjökulla Aerial Test Range, *U.S. Geological Survey*, accessed 2026.

[11] METASENSING: MetaSensing Participates to NATO-SET-250 Trials with Its Airborne SAR, *MetaSensing*, 2019.

[12] CHRZANOWSKI, K.: Testing Thermal Imagers: Practical Guidebook, Military University of Technology, Warsaw, 2010.

[13] BIJL, P., VALETON, J. M., DE JONG, A.N.: TOD Predicts Target Acquisition Performance for Staring and Scanning Thermal Imagers, *Proceedings of SPIE*, Vol. 4030, 2000.

[14] DROBNJAK, S., MILONJIĆ, Z., ĐORĐEVIĆ, D., SIMIĆ, M.: Flash flood susceptibility mapping using LiDAR-Derived DEM and machine learning algorithms: Ljubovića case study, Serbia. *Open Geosciences*, 17(1), p.20250859, 2025.

[15] PERIĆ, D., LIVADA, B., PERIĆ, M., VUJIĆ, S.: Thermal Imager Range: Predictions, Expectations, and Reality, *Sensors*, 19(15), 3313, 2019.

[16] ĐORĐEVIĆ, D., ĐURIĆ, U., BAKRAČ, S., DROBNJAK, S., RADOJČIĆ, S.: Using historical aerial photography in landslide monitoring: Umka case study, Serbia. *Land*, 11(12), p.2282, 2022.

[17] CAI, Y., ZHOU, Y., CHEN, Z., WANG, T., CHENG, X.: Review of Target Geo-Location Algorithms for Aerial Remote Sensing Cameras without Control Points, *Applied Sciences*, 12(24), 12689, 2022.

[18] ZHANG, X., LIU, C., DING, Y., ZHANG, H., XIU, J., KUANG, H.: Precise Target Geo-Location of Long-Range Oblique Reconnaissance System for UAVs, *Sensors*, 22(5), 1903, 2022.

[19] SIMIĆ, M., VAKANJAC, B., DROBNJAK, S.: NDGRI: A Novel Sentinel-2 Normalized Difference Gamma-Radiation Index for Pixel-Level Detection of Elevated Gamma Radiation. *Remote Sensing*, 17(19), p.3331, 2025.

[20] SIMIĆ, M., VAKANJAC, B., DROBNJAK, S., NOVKOVIĆ, I., JOVANOVIĆ, J.M.: Deep learning classification of elevated gamma-radiation zones using Sentinel-2 spectral data. *Open Geosciences*, 18(1), p.20251000. 2026.

[21] JOVANOVIĆ, J.M., STOJANOVIĆ, M., JANKOVIĆ, T., DROBNJAK, S., DJORDJEVIĆ, D., BANKOVIĆ, R., RADOVANOVIĆ, M., ŽIVKOVIĆ, L., GAJIĆ, T., DEMIROVIĆ BAJRAMI, D., TRETIAKOVA, T.N.: A case study on the Danube Limes in Serbia: valorisation and cartographic analyses of selected tourism products. *Sustainability*, 14(3), p.1480, 2022.

[22] DROBNJAK, S., SEKULOVIĆ, D., AMOVIĆ, M., GIGOVIĆ, L., REGODIĆ, M.: Central geospatial database analysis of the quality of road infrastructure data. *Geodetski vestnik*, 60(2), pp.270-284, 2016.

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

NOVEL STRATEGIC TRENDS IN WEAPON SYSTEM ACQUISITION

NEBOJŠA NIKOLIĆ
Faculty of Project and Innovation Management "Prof. dr Petar Jovanović", Belgrade, nebojsa.nikolic@pmc.edu.rs

MILOŠ STOJANOVIĆ
Faculty of Project and Innovation Management "Prof. dr Petar Jovanović", Belgrade, milos.stojanovic@pmc.edu.rs

MIROSLAV MITROVIĆ
Faculty of Project and Innovation Management "Prof. dr Petar Jovanović", Belgrade, miroslav.mitrovic@pmc.edu.rs

Abstract: *This paper presents contemporary trends in a weapon system acquisition programs through analyses of the newest strategic document related to the subject in few selected but relevant cases – the U.S. and the EU. Observed changes in conducting modern warfare are perceived as radical changes which generate a need for change of strategic documents and approaches in many fields, including ackuisition of weapons and military equipment. Main characteristics of those changes are related to rapid development and fast fielding of new technological solutions and systems, innovative approaches, application of artificial intelligence and advanced information and communication technologies.*

Keywords: *acquisition, strategy, defense industry, weapon systems, program management.*

1. INTRODUCTION

Contemporary armed conflicts during past several years are characterized by huge changes at all levels – tactical, operational and strategic. Changes are not linear nor straight forward, but iterative and multidirectional with mutual influence and impact. Dynamics of involvement of new weapon systems and combat use of dual-use technologies, became very high. The result of that, is enlarged pressure on weapon system suppliers to obtain a better, and better weapons, in a shorter time, and in required quantities.

At the opposite ends of supply chain for weapon systems and technologies, are producers (weapons factories) and end users (military units). In between the two, are organizational entities which are the main bridge between users of weapons systems and the producers. That intermediary is called acquisition system. Certainly, in various countries and in various historical time acquisition system has various names and organizational location inside the structure of armed forces and/or ministry of defense.

Principally, the main task of acquisition system is to obtain exactly that what is needed and required by the end users. However, that task is very complex, demanding and uncertain. In a case of civilian supply chains, the role of main mediator between producers and users is delegated to the free market. This fact is one of key differences between military and civilian supply chains. In that sense, defense acquisition system has very important role as main bridge between production side with defense industry as the cornerstone, and the armed forces as end-users.

Studying topics similar to this one is useful from the perspective of drawing lessons from foreign experiences and possible application for one's own needs. Also, knowledge on acquisition system and defense industrial base of the U.S. could be very useful for foreign parties which are interested in technological cooperation and arms-trade with industrial and technological companies as well as with defense institutions in the U.S.

Next sections of this paper present in short main points from the newest acquisition strategy of the U.S. Department of War (former Department of Defense), issued in 2025. Then follows presentation of novel trends in the EU and its specific characteristics. Fourth section offers two illustrative examples: first, review of growing investment in acquisition in NATO countries; and second, basic information for one of the newest, the largest and the most expensive complex acquisition program in defense, known aa "Golden Dome".

2. NEW ACQUISITION STRATEGY OF THE U.S.

New Acquisition strategy of the U.S. was issued by the end of last year from the U.S. Department of War [1]. It was a result of implementation of one executive order of the President of the U.S., [1], issued in the April, the same year, which was related to modernization of defense acquisition and to defense industry.

Initial assumption in new acquisition strategy points out two moments. First, existing acquisition system is perceived as outdated ("antiquated"). Second, defense industrial base is considered as insufficient and limited in capacity and performance ("atrophied"). Those two

Corresponding author: N. Nikolić;
E-mail: nebojsa.nikolic@pmc.edu.rs
DOI: 10.5937/OTEH26158N

positions combined with the huge changes in modern armed conflicts and consequently requirements for enhanced capabilities of military forces, were more than sufficient causes for creation of new acquisition strategy.

2.1. Main goals of new acquisition strategy

Main strategic goals of the new acquisition strategy, were defined along three directions as follows. First, implementation of new technologies in order to modernize systems at the level which is higher than it could be the case with potential adversaries. Second, enlargement production capacity of defense industry with the focus on those who produce weapons, ammunitions and key systems. Third, transform acquisition system and industrial base in such a way that it corresponds to a wartime regime and is capable to respond to the urgent needs.

First two strategic directions are more general and logical having in mind that modernization is present and relevant for each field of industry and business, while enlargement of industrial production of military equipment could be seen as a response to a global weapons market trends and increased demands. However, the third strategic goal has the largest transformational change potential because it declares the need, will and intention to transit complete defense industrial base and the acquisition system from peacetime regime towards wartime regime.

Strategic importance of defense industrial base is additionally confirmed by the newest National Defense Strategy [3], where it is defined as one of four main strategic goals for national defense (other three goals are: defend the country; deter main adversaries; and increase burden-sharing with allies and partners). This strategy says that it will be undertake “urgent actions” in order to “mobilize, renew and secure” defense industrial base.

2.2. Specific tasks of new acquisition strategy

New American acquisition strategy is intended to achieve above mention main strategic goals through defining and realization of five sets of specific tasks as follows.

First, rebuilding, or supercharging and overhaul of defense industrial base. This rebuilding efforts encompasses following actions:

- Expand number of companies which will produce military equipment and relax their joining; welcome better ideas; promote fast contracting and speed delivering of results.
- Stabilization of demands dynamics through longer and bigger deals with companies which produce weapons and military equipment. It is expected that such companies will be more willing to invest in their capacity for weapons and equipment production.
- Attract private capital investment in defense industry. This measure will impact both above actions (expanding capacity, and enlarging number of companies).
- Use off-the-shelf purchase of products, parts and services whenever it is more costly-efficient and faster.
- Renovate bid protest procedures in such a way that they are faster in a whole and more costly for those protesters who lose.
- Make contracting more flexible so that process is faster, that is achieved in a shorter time.
- Increase and systemize communication between Department of War and companies in defense industrial base, through establishing appropriate consortium.

Second, improve capacity and capability of acquisition workforce and its authorization level, through following actions:

- Attract top talents from industry and government, hire them in acquisition bodies and let them to make work better.
- Educate its own workforce through systemic advancement of the Warfighting Acquisition University.
- Sustain and keep best leaders and teams.
- Authorize acquisition program leaders to conduct full control of program outcomes quality, budget and deadlines.
- Improve capability portfolio management through optimization of purchasing of similar products in a sense efficiency, effectiveness and timely realization.
- Include acquisition managers in early phases of warfighting requirements defining.
- Increase speed of reform of the Planning, Programming, Budgeting and Execution processes, by focusing on enabling leaders’ agility to schedule financial resources, preferencing new technologies and simplify approval procedures.
- Finally, improve stimulating system of workforce by connecting their results in acquisition processes with their payment.

Third, enlarge flexibility in all phases of acquisition processes. This should result in reduced regulations and reduced time parameters for acquisitions. To achieve that, following should be undertaken:

- Digitalization of acquisition processes, smart use of Artificial Intelligence (AI) for reduction of administrative issues and planning with a main goal to reduce overall acquisition time and make it faster.
- Permanent control of acquisition programs through access to relevant data in a real-time, measure progress and make appropriate corrections if needed.
- Modernization of testing processes and infrastructure, with intention to make testing decisioning faster.
- Reducing number of bureaucratic testing procedures in favor of real testing.
- Increase quality of acquisition of information technology, particularly software and hardware.
- Reduce and replace procedure known as “Analysis of alternatives” which is demanding and time-consuming.
- Make better Cost estimation, use realistic data, make accurate prediction about real and overall cost of acquiring weapon system or equipment.
- Make transition towards commercially used accounting system.

- Rebuild Foreign military sales, strengthen industrial base capacity for export, and make effective and precise procedures to support arming allies and partners with American weapons.

Fourth, rise quality standards and technical excellence in order to produce systems of high performances, through following directions:

- Modernization of system engineering, upgrading quality standards for weapon systems, and at the same time make it more efficient and faster.
- Use open system architecture in the sense of wider use of common components and standards in weapon systems, which will enable fair compete among more companies.
- Enable adaptive acquisition framework in order to obtain fact implementation of the newest technological advancements in weapon systems.
- Re-engineering of the "kill-chains" studies, that is, reconsidering tactical procedures and decision making related to use of weapons (in short: sense-identify-track-decide-act-report chain).

Fifth, advance of lifecycle risk management. It should be realized by actions as follows:

- Size and sustain complete maintenance and repair of weapons system, by own capacities (military and government, not commercial). This intention is interesting, by the way, because there were efforts in the opposite direction [4].
- Impose comprehensive and overall control on supply chain for defense industry, and fix broken and weak points.
- Make lifecycle management as long-range sustainable, cost-efficient and with mapped risks and their possible impact.

In summary, the Department of War of the U.S. is main initiator, strategic planner and supervisor of transformation of the whole system of defense acquisition. The change dynamic is very high starting in 2025 with the presidential strategic directives, which was followed by transformation strategy in the same year, with precise and relatively short deadlines for some singular steps measured by months, and with good prospects to be realized in foreseeable future. It will be useful to follow defense acquisition transformation activities in the U.S., in the coming years and register how it advance and what problems appears in this transformation.

3. DEFENSE PROCUREMENT IN THE EU

Question of defense acquisition in the European Union and in its' member states, is also very actual and is subject of research analyses, political considerations, government decisions and industrial preparations in a longer period [5]. However, the situation here is much more complicated than it is in the case of the U.S. In the European context we have individual countries in Europe and some of them are member countries of the EU, some of them are not; some of them are members of the NATO, some are not; also, there are institutional bodies and agencies of the EU, and finally there are commercial companies in the defense industrial sector. Also, there are defense industry companies which are still owned by government.

Recent research study about European defense issues, in a separate section dedicated to defense procurement (European synonym for acquisition) point out the state-of-the-art in few concise notes, [6]:

- There are evident changes in defense procurement in European countries, where increased spending on new weapons and equipment is the main change. Also, there are some changes in approaches to defense procurement.
- There is evident trend of orientation of European states towards European based procurement, that is, buying defense gods from European producers. Not so far in the past such practice would be convicted as discrimination or at least as non-fair market behavior, but today it is accepted as politically correct and even very positive.
- In order to keep and sustain increased level of investment in defense, it will be needed reliable and permanent support from high political level. This is important for European countries due to their reluctant position towards military spending in a long period from the end of the Cold War up to 2014 and even up to 2022 (years of conflict escalation between Ukraine and Russia).
- Due to complex issues in relations among national governments and political bodies of the European Union, there are two almost simultaneous processes of a collaboration at international level inside the EU, and a fragmentation as a dealing with defense issues at internal national level.
- At the moment, in Europe, dominant approach in defense procurement is nationally focused which means prioritization of national priorities, national industrial politic, and national legislation. Consequently, preference of national approach has negative impact on unionist or collaborative approach.

It is clear that a "renaissance" of the European defense issues including but not limited only on industry, is motivated by dominant security perceptions of Russia as an adversary, particularly after beginning of open armed conflict with Ukraine in 2022. Not to forget, besides the so called "Russian threat", there was occasional critique of European countries by administration of the U.S. for low level of defense spending among European countries. Europeans were steadily resistant to that American pressure for greater defense spending, but after 2022 situation changed dramatically.

As a result, there are many changes in strategic documents, development and application of all kind of sanction, search for alternative sources of energy, transportation routes, preparation of domestic population for potential conflict, enlargement of defense budgets, re-introduction of mandatory military service, and many other measures.

European countries were main beneficiaries of the "peace

dividend" after the end of the Cold war and theirs armed forces and defense industry had not priority in national development plans. However, in a relatively short time situation has changed and instead of "reduction" and "cut-off" of armed forces and defense industrial capacities, geopolitical environment forced changes in the opposite direction -towards enlargement and growth.

3.1. European defense industry transformation strategy

There are systematic efforts at the EU level and among member states to find the way for much more collaboration, cooperation and joint efforts for common approach in defense procurement. One result of that common efforts is a novel strategic document related to the roadmap for defense industry transformation in the European Union [7].

European defense industry strategy points out at the very beginning imperative significance of innovation, speed, agility and risk-taking aspects, which is similar with the U.S. point of view. Also, the newest lessons learned from ongoing armed conflicts, like war in Ukraine, has prominent place in perceiving the requirements needed for future weapon systems.

In order to speed up involvement of new technological achievements and innovations in a new weapons development, European administration and member states have started with stronger support to technological startups in defense industry business. In that sense, in a just few years after 2022, more than 230 small companies gained support for their programs in defense related technologies, [7], such as combat application of artificial intelligence, cyber tools, use of space-based systems, drone and counter-drone systems, radar and other sensing systems, dual-use technologies and systems, etc.

3.2. European Defense Agency and procurement issue

Main strategic document of the European Defense Agency (EDA), defines five core tasks. One of those core tasks clearly confirms importance of procurement and main approach which will be used for that. It is declared as: "Aggregating demand towards joint procurement", [8].

However, in other core tasks of EDA, there are elements and aspects related to procurement, such as follows [8]: "Enabling collaborative defense research, technology and innovation"; "harmonize requirements and ensuring interoperability".

European Defense Agency is positioned as the main intergovernmental body at the EU level which is authorized and dedicated to all aspects of defense industrial and research issues. As such, EDA mission is to support national ministers of defense "along all the steps of the capability development cycle and beyond that" [8].

While EDA operational budget is at the level of several tenths of million EUR (about 68 million EUR), total defense spending in the EU is much more (more than 400 billion EUR). So, the potential for influence, at least indirect, by EDA on the European Union defense spending is huge. This fact could be motivational factor for small states to establish and sustain good relations and connections with EDA and make some cooperation in the field of defense research and defense industry and so enter in big European defense market.

4. SOME INTERESTING EXAMPLES RELATED TO ACQUISITION

4.1. Budgeting acquisition – NATO case

Budget financing of acquisition of weapon systems and equipment have had steadily growth during last ten years in many countries. It was the consequence of crucial changes in geopolitical environment, strategic shifts in global interest of great powers, and of course because of several inter-state armed conflicts and innovative application of new technologies in those conflicts.

Scale of change in share of weapons and equipment expenditure inside the defense budget was tremendous. Very illustrative are data for NATO countries [9]:

- In year 2014., only several countries have had expenditure for weapons and equipment more than 20% of the defense budget. Those were: the U.S., the U.K., Sweden, Norway, France, Turkey, Estonia and Luxemburg [9].
- However, in year 2025, all NATO countries, except Belgium, reached the level of 20% of equipment expenditure.

Additionally, that grows of spending money on new weapon and equipment were extremely high and much more than 20% – meaning many times more. Here are some examples of that grow:

- Poland spent more than 54% in 2025, compared with below 20% in 2014.
- Finland spent more than 53% in 2025, compared with below 15% in 2014.
- Hungary, Latvia, Greece, Montenegro, North Macedonia and Czeckia spent more than 30% in 2025, compared with below 10% in 2014 [9].

4.2. Agile acquisition for the Golden Dome

The newest acquisition initiative for ballistic missile defense in the U.S., is known under a name: Golden Dome. The name is chosen as association on the Israeli's Iron Dome – an already fielded system for ballistic missile and aerial defense which has proved itself as efficient during last several years.

As official start of the Golden Dome program could be taken January 2025, when President of the U.S. issued executive order for development of this acquisition program [10]. Vision of the finishing of this program, is projected up to the end of the mandate of current presidential administration, which is in 2029.

Cost estimation for the program are relatively unprecise and vary from $175B up to $1,300B. Lower number is more related to estimation of the development of all subprograms and prototyping, while upper cost estimation

relates more on adding operational and logistical costs for additional 20 years of service.

Acquisition of the Golden Dome program some authors called as "non-traditional" due to its unique characteristics such as follows:

- The highest strategic importance – it is intended to protect the whole territory of the U.S. from all kinds of aerial and space threats (ballistic missiles, hypersonic vehicles of all kind, etc).
- Global deployment - in order to fulfil its mission, it has to be deployed globally including the space.
- Complexity – it will consist of many kind of systems and platforms in a large number (hundreds of space based satellites, latter thousands for surveillance, tracking, guiding; layered missile systems on earth, surface, air and even in space; combat information network, command and control centers and infrastructure; logistical sustainment, etc).
- Enormous cost of the program even with the lower estimate.
- Technical aspects of development of all subprograms and projects – many technical solutions and system are yet to be developed.
- Relatively short time for an such a large, innovative and complex program.

5. CONCLUSION

From the point of view of strategic management, huge changes in a "military business" which are evident from ongoing armed conflicts around the globe, present fundamental changes in a wider environment of comprehensive defense sector perceived as a strategic organization. In such cases, the organization have to react to that changes in its environment, and creation of new strategy and strategic plans are starting steps. New "Acquisition Transformation Strategy" of the U.S. is exactly such case, as well as corresponding strategic documents in the European Union.

Quality and successfulness of each strategy will be seen after some period and by the results of its implementation in practice. Future research should be oriented in that direction - tracking successfulness of application of strategies and strategic plans, and analysis of problems in its realization.

In a wider sense of acquisition goals, future research could be focused as well, towards search for still unresolved problems in conducting modern combat and defense, such as:

- Ballistic missile defense;
- Counter-drone technologies, systems and tactics;
- Radar and sensing technologies;
- Autonomuos swarming drones;
- Comand & Control systems for faster and more reliable decision making at all levels.

It is interesting that some of above challenges has longer history, that is, still with no solutions during several decades. Such case is with the long-standing problem of ballistic missile defense. Namely, this problem in real combat conditions was identified long ago, during the First Gulf War, in 1991. In that time a wide global coalition lead by the U.S., and authorized by the U.N. mandate, attacked Iraq forces in Kuwait, forcing them to leave Kuwait which they were occupied in a year before. In a after-action-reviews and in comprehensive analyses after the war, military experts and war-participants of the highest level, concluded that ballistic missile defense is a huge and unresolved problem, [11]. This problem is still actual and search for effective and efficient solution continues.

References

[1] The U.S. Department of War: *Acquisition Transformation Strategy – Rebuilding the Arsenal of Freedom*, 2025.

[2] The White House: *Modernizing Defense Acquisition and Spurring Innovation in the Defense Industrial Base*, Executive Order 14265 of April 9, 2025, of the President of the U.S., 2025.

[3] The U.S. Department of War: *National Defense Strategy – Restoring peace through strength for a new golden age of America*, 2026.

[4] Nikolic,N., Milinovic,M., Jeremic,O., Jankovic,R., Jezdimirovic,M.: "Some problems of field maintenance in the Stryker type brigades", *Proceedings, 5th International scientific conference on defensive technologies, OTEH 2012*, 18-19 Sept. 2012, Belgrade, 756-759.

[5] Nikolic,N., Milinovic,M., Jankovic,R.: "Strategic Framework for European Defence Research, Technology and Industry**",** *Proceedings 13th International Symposium Sym-Org-2012*, 05-09 June 2012, Zlatibor, Serbia, 1743-1747.

[6] The International Institute for Strategic Studies (IISS): *Progress and Shortfalls in Europe's Defense: An Assessment*, 2025.

[7] European Commission: *EU Defense Industry Transformation Roadmap*, 2025.

[8] European Defense Agency: *2024 Long Term Review*, 2024.

[9] NATO, Public Diplomacy Division: *Defense Expenditure of NATO countries (2014-2025),* 2025.

[10] Poskey, D., Sulak, K.: "Nontraditional Acquisition--Making the Golden Dome for America A Reality", *Defense Acquisition*, *54*(303), 20-26, 2025.

[11] Mitrović, M., Nikolić, N.: *Hibridni rat: Doprinos definisanju koncepta, sadržaja i modela delovanja*. Medija centar "Odbrana", 2022. (book title in English: *Hybrid War: Contribution to Defining the Concept, Content and Model of Action*)

12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

STANDARDIZATION AS A REGULATOR FOR THE USE OF ARTIFICIAL INTELLIGENCE IN FRAUD DETECTION

IGOR ĐORIĆ
Military Tehnical Institute, Belgrade, igor.djoric@mod.gov.rs

SNEŽANA ZUROVAC
Military Tehnical Institute, Belgrade, zurovac@medianis.net

Abstract: *Artificial intelligence (AI) is becoming one of the most significant drivers of the digital transformation of the financial sector, enabling the development of advanced systems for the detection and prevention of financial fraud. By applying machine algorithms learning, deep learning and analysis of large data sets, it is possible to identify suspicious behavior patterns in real time and significantly improve the effectiveness of risk management systems. However, the growing application of artificial intelligence simultaneously raises issues of transparency of algorithms, data protection, responsibility for decisions made, ethical aspects and cyber security. Therefore, standardization represents one of the key regulatory mechanisms that enables the responsible, safe and reliable application of systems based on artificial intelligence. The aim of this review is to analyze the role of standardization in regulating the application of artificial intelligence in the detection of financial fraud. Special attention is paid to ISO/IEC international standards, the European Act on artificial intelligence (EU AI Act), the NIST framework for artificial intelligence risk management, as well as the challenges brought by the abuse of artificial intelligence, cybercrime and new forms of digital fraud. The paper synthesizes contemporary scientific knowledge and points to the importance of integrating technical standards and regulatory frameworks in order to improve the security of the financial system and increase user confidence in AI technologies...*

Keywords: *artificial intelligence, standardization, financial fraud, financial sector, risk management.*

1. INTRODUCTION

Digitization of financial services represents one of the most significant processes of modern economic development. Electronic banking, mobile payments, cryptocurrencies, open banking and digital identity have enabled faster, more efficient and more accessible financial transactions, on the one hand. In contrast, but at the same time, new information technologies have enabled criminal groups to apply sophisticated attack methods that exceed the capabilities of traditional protection systems.

Today, financial institutions on a global level process billions of transactions every day, which makes manual control an almost impossible task. In response to this challenge, artificial intelligence-based systems have been developed that enable automatic recognition of anomalies, analysis of user behavior patterns, and assessment of the probability that a particular transaction represents a fraud attempt. Compared to traditional systems based on predefined rules, machine learning algorithms can continuously learn from new data, adapt to changes in user behavior and identify previously unknown patterns of behavior, which can indicate possible fraudulent activities.

lthough such systems significantly increase the efficiency of financial institutions, their implementation raises numerous questions. One of the biggest challenges is the so-called "black box" problem, i.e. the inability to clearly explain how complex algorithms arrive at a certain decision. Lack of transparency can lead to wrong risk assessments, discrimination of certain categories of users or unjustified blocking of financial transactions. In addition, artificial intelligence systems can be the target of cyber attacks or the subject of abuse by criminal organizations.

Contemporary research shows that the business value of artificial intelligence does not depend solely on the quality of algorithms, but also on how organizations manage data, risks and regulatory requirements. The sustainable implementation of AI requires the integration of technological, organizational and management aspects, with an appropriate regulatory framework being one of the key prerequisites for creating business value (Enholm, I.M., et al., 2022).

t is precisely for this reason that standardization gains strategic importance. Standards enable the establishment of unique rules for the design, development, implementation and maintenance of artificial intelligence systems. Their application contributes to increasing interoperability, transparency, data management quality and user trust. Standardization has a significant positive impact on economic development because it encourages innovation, increases productivity, and reduces the cost of implementing new technologies (Blind, K., et al., 2011). This conclusion is especially important in the field of

Corresponding author: I. Đorić ;
E-mail: igor.djoric@mod.gov.rs
DOI: 10.5937/OTEH26159D

artificial intelligence, where standards are the basis for aligning technological development with regulatory requirements.

n addition to economic aspects, standardization plays an important role in preserving the security of the digital environment. Modern AI systems can be misused to create fake digital content, automate cyber attacks or carry out sophisticated financial fraud. The development of deepfake technologies, generative artificial intelligence and automated systems for social engineering shows that the same technology can have both positive and negative effects. Okolie points to the serious challenges that deepfake technologies pose for the privacy and security of individuals, while Montasari emphasizes that the abuse of artificial intelligence is becoming one of the key issues of national and cyber security. (Okolie, C., 2023) (Montasari, R., 2022)

An additional challenge is the fact that criminal and extremist organizations are increasingly using artificial intelligence to automate propaganda, create fake content and plan attacks. The development of generative artificial intelligence significantly increases the possibilities of misuse of modern information technologies for extremist and terrorist purposes (Đorić, M., & Milošević, T., 2021).

he development of international standards is one of the most effective mechanisms for limiting the criminal use of AI systems (Ayyaswamy, K., et al., 2025). In this context, standardization is not only a technical issue but also a regulatory instrument that defines the requirements for security, transparency, responsibility and ethical application of artificial intelligence systems. In recent years, numerous international standards have been developed, among which ISO/IEC 42001, ISO/IEC 23894 and ISO/IEC 38507 stand out, while the European Union, by adopting the AI Act, laid the foundations for the first comprehensive regulatory framework for the application of artificial intelligence.

Starting from the mentioned facts, the aim of this paper is to analyze the role of standardization as a regulatory mechanism for the responsible application of artificial intelligence in the detection of financial fraud. Through a review of modern literature, the technological aspects of AI application, the importance of international standards, the regulatory framework of the European Union, as well as the security challenges arising from the increasingly intensive use of intelligent systems will be analyzed.

2. ARTIFICIAL INTELLIGENCE IN THE FINANCIAL SECTOR

Today, artificial intelligence represents one of the key technologies of the digital transformation of the financial sector. Its application includes almost all business segments of financial institutions - from the automation of customer support and credit risk assessment to investment portfolio management, fraud detection and money laundering prevention. The development of advanced algorithms enabled the transition from traditional rule-based systems to adaptive models that learn from data and continuously improve their performance.

The banking sector was among the first to recognize the potential of artificial intelligence. Modern banks use AI to process loan applications, assess customer solvency, personalize financial services, optimize business processes and improve customer experience. In addition, machine learning algorithms are used to identify suspicious transactions in real time, thereby significantly reducing financial losses caused by fraud.

The business value of AI results from a combination of technological capabilities, organizational capacity and quality management. AI contributes to increasing operational efficiency, improving decision-making processes and developing innovative business models, but only if its implementation is accompanied by appropriate management mechanisms and a regulatory framework (Enholm, I.M., et al., 2022).

he financial sector represents a particularly sensitive environment for the application of AI due to the large volume of data at its disposal and the high level of regulatory requirements. Banks and other financial institutions process huge amounts of personal, financial and biometric data, which is why artificial intelligence systems must ensure a high level of accuracy, reliability and privacy protection. Any wrong assessment can lead to financial damage, damage to the reputation of the institution or violation of the user's rights.

herefore, in the last few years, the need for responsible application of artificial intelligence (Responsible AI) has been increasingly emphasized, which implies transparency of algorithms, the possibility of explaining the decisions made, data protection, model fairness and continuous risk management. The following chapter will analyze the application of these technologies in the detection of financial fraud, as one of the most important areas of their use.

3. ARTIFICIAL INTELLIGENCE IN DETECTING FINANCIAL FRAUD

inancial fraud is one of the biggest challenges of the modern financial system. The global digitization of banking and payment services has led to a significant increase in the number of electronic transactions, but at the same time it has enabled the development of new forms of criminal activities. Traditional fraud detection systems, based on predefined rules and manually created risk indicators, have shown limited effectiveness in the face of increasingly complex and dynamic patterns of criminal behavior. In response to these challenges, financial institutions are increasingly applying artificial intelligence and machine learning methods to improve the processes of identifying, analyzing and preventing financial fraud.

he application of artificial intelligence enables the processing of large amounts of structured and unstructured data in real time. Unlike traditional systems, AI models do not depend exclusively on pre-defined rules, but independently identify patterns of behavior that indicate potential abuses. This approach is particularly important in situations where criminals continuously

change their methods of committing fraud in order to avoid existing protection mechanisms.

In the literature, financial fraud is most often classified as payment card fraud, electronic banking fraud, identity theft, money laundering, insurance fraud, investment fraud, tax fraud, and digital currency fraud. Each of the mentioned categories is characterized by specific patterns of behavior and different levels of risk, which is why artificial intelligence systems must be adapted to the specific business environment.

One of the most important areas of application of AI is fraud detection in payment transactions. During the processing of card transactions, algorithms analyze a large number of parameters, including the amount of the transaction, the geographical location of the user, the execution time, the frequency of purchases, the device from which the transaction was executed, the IP address and the previous behavior of the user. By combining this data, it is possible to estimate in a very short time the probability that a particular transaction represents a fraud attempt.

The application of artificial intelligence enables the processing of large volumes of structured and unstructured data in real time. Unlike traditional systems, AI models do not rely only on predefined rules, but can independently identify behavioral patterns that may indicate potential, fraudulent activities.

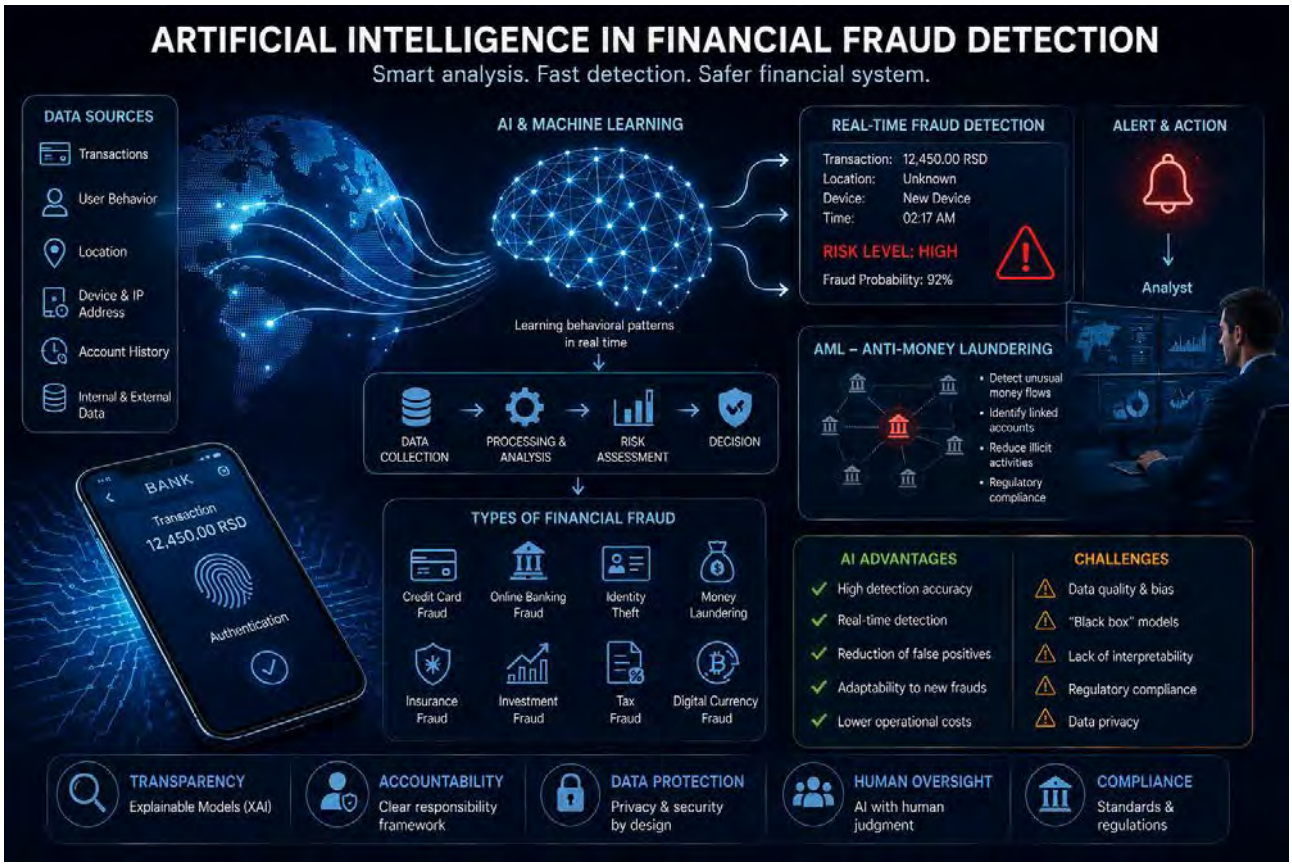


Figure 1. Conceptual illustration of the application of artificial intelligence in financial fraud detection

In addition to card fraud, an important area of application is the prevention of money laundering (Anti-Money Laundering - AML). Financial institutions analyze millions of transactions every day to identify unusual cash flows, linked accounts and potentially illegal financial activity. By applying machine learning algorithms, it is possible to identify patterns that would go unnoticed by traditional methods, thereby increasing the efficiency of regulatory controls and reducing business costs.

Methods based on machine learning achieve significantly better results than classical statistical models, especially in cases where large and unbalanced data sets are involved (Phua, C., et al., 2010). Similarly, adaptive systems represent the most effective approach for detecting new patterns of financial fraud, as they allow continuous adaptation to changes in perpetrator behavior (Bolton, R. J., & Hand, D. J., 2002).

In addition to increasing detection accuracy, artificial intelligence contributes to reducing the number of false positive results. In traditional systems, a large number of legitimate transactions are wrongly marked as suspicious, leading to user dissatisfaction and increased operational costs. By applying modern AI models, it is possible to achieve a better balance between the sensitivity of the system and the number of false warnings.

However, despite the significant benefits, the

implementation of artificial intelligence in the financial sector faces numerous challenges. The quality of the models depends largely on the quality of the data used to train them. Incomplete, outdated or biased data can lead to wrong risk assessments and discriminatory decisions. An additional problem is the fact that many deep learning models behave like a "black box", which is why users and regulators are often unable to understand the decision-making process.

This is precisely why modern regulatory frameworks increasingly insist on the development of explainable artificial intelligence (Explainable Artificial Intelligence - XAI), which enables the interpretation of the results of algorithms and increases user confidence in automated decision-making systems. The transparency of the model is one of the basic prerequisites for their application in the financial sector, where every decision can have significant legal and economic consequences.

4. METHODS AND ALGORITHMS OF ARTIFICIAL INTELLIGENCE IN THE DETECTION OF FINANCIAL FRAUD

The development of machine learning methods has enabled the application of a large number of algorithms in financial fraud detection systems. The choice of the appropriate algorithm depends on the available data, the type of fraud, the accuracy requirements of the model and the possibility of interpreting the results.

The most commonly used approach is supervised learning, where the algorithm is trained using historical data that contains information about whether a particular transaction was legitimate or fraudulent. Among the most commonly applied algorithms are logistic regression, Decision Trees, Random Forest, Support Vector Machine, Gradient Boosting and XGBoost. These models achieve high accuracy in the classification of transactions and enable relatively simple integration into existing banking systems.

When historical data is not available or is incomplete, unsupervised learning methods are applied. Their goal is not to classify pre-known patterns but to identify anomalies that deviate from the usual behavior of users. This group includes Clustering Algorithms, Principal Component Analysis (PCA), Isolation Forest and various anomaly detection models. Such approaches are particularly important for the detection of new forms of fraud that have not been registered before.

In recent years, deep learning models have been increasingly used. Neural networks enable the analysis of very complex relationships between a large number of variables and achieve extremely good results in the processing of unstructured data. Recurrent neural networks (RNN) and Long Short-Term Memory (LSTM) models are particularly suitable for the analysis of time series of transactions, while convolutional neural networks (CNN) find application in the analysis of digital documents, biometric data and identification documents.

An increasingly important role is also played by Graph Neural Networks (GNN), which enable the analysis of relationships between users, bank accounts and financial transactions. Unlike classical algorithms, GNN models analyze the network of connections between different actors and enable the identification of organized criminal groups and complex money laundering schemes.

In addition to machine learning algorithms, financial institutions also apply Natural Language Processing (NLP). NLP models are used to analyze email, user communications, suspicious activity reports and text documents. Application of large language models enables automated identification of attempted phishing attacks, fraudulent payment requests and other forms of social engineering.

Although the mentioned algorithms achieve exceptional results, their application requires appropriate infrastructure, high data quality and continuous monitoring of model performance. The business value of AI does not derive solely from the choice of algorithm, but from the organization's ability to integrate technology solutions with business processes, data management and regulatory requirements (Enholm, I.M., et al., 2022).

In addition to technical challenges, special attention should be paid to the ethical aspects of the application of algorithms. Models trained on biased data sets can lead to discriminatory decisions or unwarranted labeling of certain groups of users as high risk. Therefore, modern financial fraud detection systems must be developed in accordance with the principles of responsible artificial intelligence, which include transparency, fairness, responsibility and privacy protectioni.

All of the above indicates that the effectiveness of financial fraud detection systems no longer depends only on the development of advanced algorithms, but also on the existence of appropriate standards that regulate the way they are developed, implemented and managed. This is precisely why standardization represents a necessary regulatory framework for the sustainable application of artificial intelligence in the financial sector.

4.1 Benford's law as statistical support for AI detection of financial fraud

In addition to machine learning algorithms, financial forensics also uses statistical methods that can serve as an initial filter for extracting unusual data. One of the most famous methods is Benford's law, that is, the law of the first digit, which describes the expected frequency of initial digits in naturally occurring numerical sets. Its application includes financial reports, accounting records, public spending and other large databases in which deviations from the expected distribution can represent a signal for additional verification (Đorić, I., et. al., 2021).

The importance of Benford's law for artificial intelligence systems is reflected in the possibility to use the results of digital analysis as additional features when training and applying models for anomaly detection. Deviation from the Benford distribution does not in itself prove the existence of fraud, but directs the auditor's or algorithm's attention to data that requires more detailed analysis.

Therefore, Benford's law can be seen as a complementary instrument to supervised and unsupervised machine learning models, rather than a replacement for them.

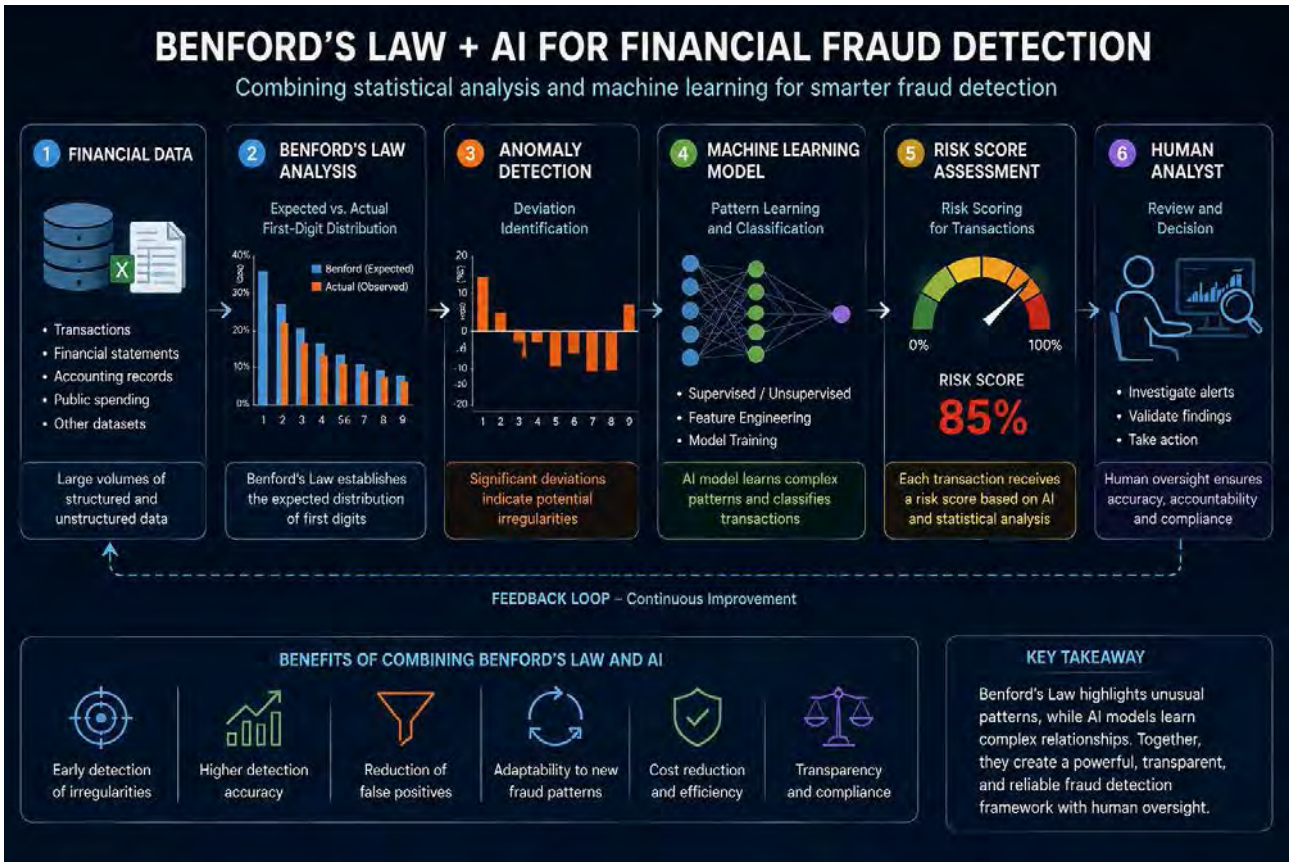


Figure 2. Benford's Law and AI-based financial fraud detection framework

Standardization of this combined methodology requires defining the conditions under which the application of Benford's law is justified, minimum data quality and volume, deviation thresholds, validation procedures and mandatory human supervision. This reduces the risk that a statistical deviation is automatically interpreted as fraud and ensures the traceability of the entire decision-making process.

5. STANDARDIZATION AS A REGULATOR OF AI APPLICATION IN FRAUD DETECTION

The development of artificial intelligence in recent years has significantly outpaced the development of regulatory frameworks that should ensure its safe and responsible application. Despite the undoubted advantages that artificial intelligence systems bring to the financial sector, their application raises numerous questions related to the transparency of algorithms, responsibility for decisions made, data protection, cyber security and ethical aspects of automated decision-making. For this reason, standardization becomes one of the most important instruments for establishing a balance between technological innovation and regulatory requirements.

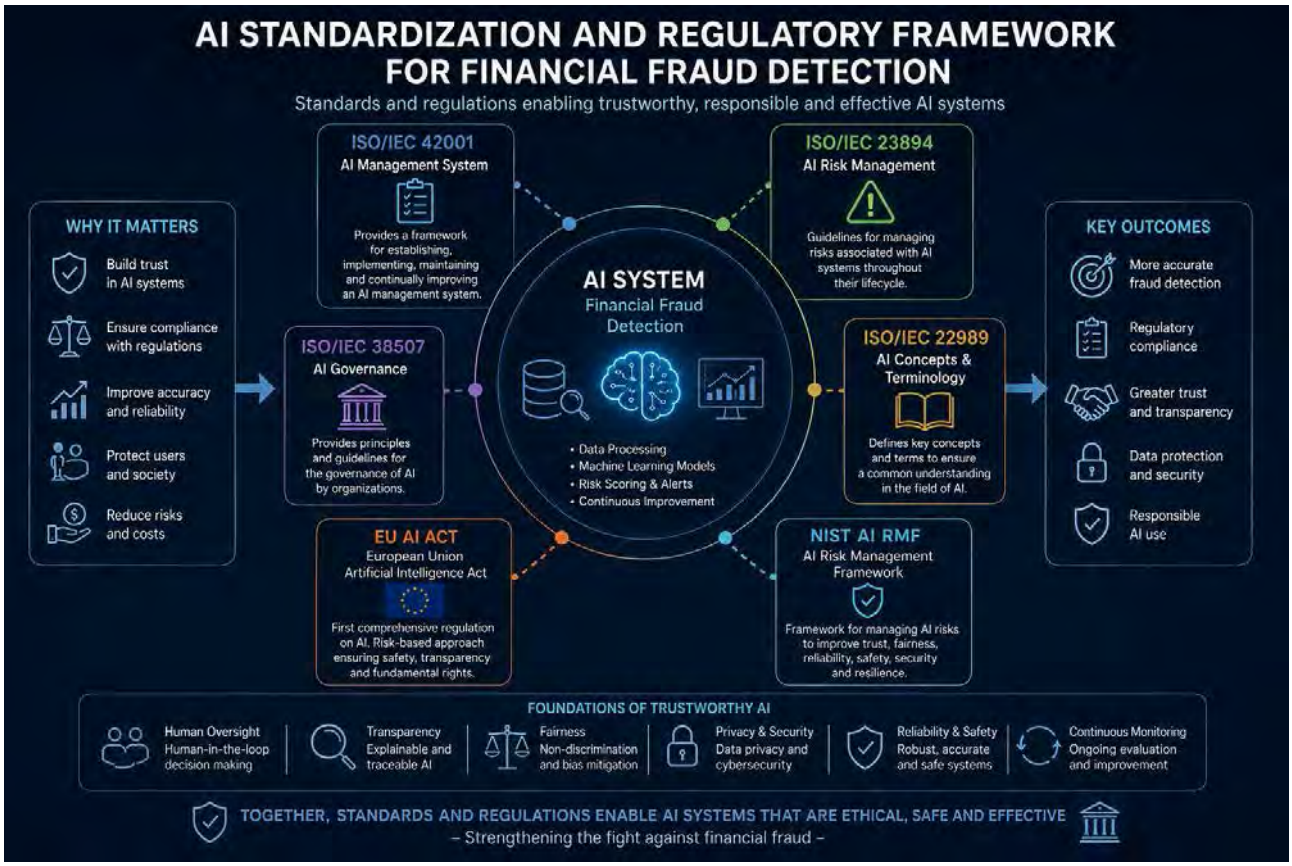


Figure 3. AI Standardization and Regulatory Framework for Financial Fraud Detection

Standardization represents the process of development and application of unique technical, organizational and management rules that enable interoperability, security and reliability of information systems. Standards have a significant positive impact on the national economy as they increase productivity, encourage innovation, reduce development costs and facilitate international trade. Standardization is not an administrative burden, but an important factor in economic development and technological progress. From there, standards represent the basis for building trust between technology producers, regulators and end users, so this conclusion gains importance in the field of artificial intelligence application (Blind, K., et al., 2011).

The field of artificial intelligence is in a phase of intensive development, which is why standardization must follow technological innovations in order not to represent a limitation for the development of new solutions. Standards should be flexible enough to support innovation, yet precise enough to enable safe and responsible deployment of AI systems. (Zielke, T., 2020). This approach is particularly important in the financial sector, where algorithms make decisions every day that have a direct impact on the financial resources of citizens and business entities.

The special importance of standardization is reflected in increasing the transparency of algorithms. Financial institutions must be able to explain to regulatory authorities why a particular transaction has been flagged as suspicious or why a particular customer has been classified as high-risk. In this sense, the standards represent the basis for the development of explainable artificial intelligence (Explainable AI - XAI), which enables the interpretation of the work of algorithms and increases user confidence in automated decision-making systems.

5.1 SO/IEC 42001 - Artificial intelligence management system

y publishing the ISO/IEC 42001:2023 standard, international standardization made a significant step forward in regulating the application of artificial intelligence. It is the first international standard that defines the requirements for establishing an Artificial Intelligence Management System (AIMS).

he structure of the standard is based on the principles already known from the ISO 9001 and ISO/IEC 27001 standards, but it is adapted to the specifics of the AI system. The standard includes planning, development, implementation, performance monitoring, risk assessment and continuous improvement of artificial intelligence systems.

or financial institutions, ISO/IEC 42001 is of particular importance because it enables the establishment of clear organizational responsibility for the development and use of AI models. The standard defines the obligation to identify interested parties, assess the impact of AI systems, manage risks, document the development process and conduct regular internal checks.

he application of this standard contributes to increasing the transparency of financial fraud detection systems and

facilitates the demonstration of compliance with regulatory requirements, including the AI Act and data protection regulations.

5.2 SO/IEC 23894 – Risk management of AI systems

One of the most important challenges in the application of artificial intelligence is risk management. Unlike traditional information systems, AI models can change their behavior over time due to new data or additional model training.

ISO/IEC 23894:2023 defines guidelines for the identification, assessment, treatment and continuous monitoring of risks during the entire lifecycle of an AI system.

The standard recognizes several risk categories: technical risks, organizational risks, legal risks, ethical risks, privacy risks and security risks.

In the case of financial fraud detection systems, there are particularly significant risks associated with misclassification of transactions, discrimination of users, inadequate data quality and the possibility of manipulating models through adversarial attacks.

Application of this standard enables financial institutions to continuously assess the reliability of AI systems and identify new security challenges in a timely manner.

5.3 SO/IEC 38507 and the management of organizations using AI

Unlike the ISO/IEC 42001 standard, which defines an AI management system, the ISO/IEC 38507 standard focuses on the management of organizations using artificial intelligence.

The standard emphasizes management's responsibility for: strategic decision-making, ethical application of AI, risk management, protection of user interests and compliance with laws.

In the financial sector, management is responsible for the decisions made by AI systems. ISO/IEC 38507 therefore emphasizes that no automated decision should be completely exempt from human supervision.

5.4 SO/IEC 22989 – Terminology and basic concepts

One of the frequent problems in the field of artificial intelligence is the different interpretation of basic terms. ISO/IEC 22989 establishes a uniform terminology that allows regulators, researchers and manufacturers to use the same definitions.

he unified terminology is the basis for the development of other standards and regulatory documents.

5.5 Controlling and internal auditing as elements of AI systems management

he standardization of the AI system must be connected with the existing mechanisms of internal control, controlling and internal auditing. These mechanisms ensure that the technology performance of the model is not viewed in isolation from the organization's goals, regulatory compliance, asset protection and risk management quality.

Internal control can be preventive, ongoing (detective) and subsequent (corrective). This division is directly applicable to the life cycle of artificial intelligence systems for financial fraud detection. Preventive controls include checking the origin and quality of data, approving models before commissioning and segregation of responsibilities. Detective controls include continuous monitoring of performance, changes in data distribution, false positives and false negatives, and attempts at model manipulation. Corrective controls include incident investigation, model retraining, rule changes, and documenting lessons learned (Đorić, I., et al., 2022).

Controlling is focused on the future and on improving the efficiency and effectiveness of the implementation of plans, while internal auditing provides an independent and objective assessment of the control system and the performed activities (Đorić, I., et al., 2022). In the application of AI, this means that controlling monitors whether the system contributes to reducing losses, speeding up the processing of warnings and achieving the organization's goals, while internal auditing checks the legality, reliability of data, integrity of records, documentation of decisions, adequacy of human supervision and compliance with ISO/IEC standards and regulatory requirements.

The special value of this approach is the linking of financial and non-financial indicators. In addition to the amount of prevented losses, it is necessary to monitor the accuracy and responsiveness of the model, the rate of false warnings, the time to solve cases, the number of user complaints, the stability of the model and the quality of the explanation of decisions. Thus, internal auditing and controlling become operational mechanisms for implementing the requirements of ISO/IEC 42001 and ISO/IEC 23894, and standardization gets practical application through clearly defined responsibilities, documented controls and measurable success criteria.

6. EU REGULATORY FRAMEWORK AND INTERNATIONAL GUIDELINES

In addition to international standards, the development of a regulatory framework is a key element of the responsible application of artificial intelligence. The European Union has adopted the AI Act, the first comprehensive regulation governing the development, marketing and use of artificial intelligence systems.

A key feature of the AI Act is a risk-based approach. AI systems are classified into four categories:

- systems with **unacceptable risk**,
- **high-risk** systems,
- **limited risk** systems,
- **minimal risk** systems.

Systems used to assess credit risk, detect financial fraud and prevent money laundering are classified as high-risk systems. They are subject to strict requirements that include risk management, high data quality, technical documentation, record keeping, algorithm transparency, human oversight and cyber security.

An important international framework is the NIST AI Risk Management Framework (AI RMF 1.0), which defines four basic functions of managing AI systems:

Govern - management and responsibility;
Map - risk identification;
Measure - risk assessment and measurement;
Manage - continuous management and improvement.

For financial institutions, it is particularly important that the principles of the NIST framework are compatible with the requirements of ISO/IEC 42001 and the AI Act. This enables the development of a unique management system that simultaneously meets international standards and regulatory requirements.

It can be concluded that standardization and regulation do not represent competing mechanisms, but mutually complementary instruments. While regulatory regulations define the legal obligations of organizations, international standards provide practical guidelines for their implementation. It is the synergy of these two approaches that is the basis for the responsible and sustainable application of artificial intelligence in the detection of financial fraud.

7. RISKS AND ABUSES OF ARTIFICIAL INTELLIGENCE

Although the application of artificial intelligence brings significant benefits to the financial sector, the same technology can be used to carry out various forms of criminal activities. The development of generative artificial intelligence, Large Language Models (LLM), technologies for voice and image synthesis, as well as automated systems for analyzing large amounts of data, has enabled criminal groups to develop new methods of committing financial fraud. This is precisely why the question of the security of AI systems becomes as important as the question of their functionality.

One of the most serious challenges is the emergence of deepfake technologies. Advanced models of generative artificial intelligence make it possible to create extremely convincing videos, audio recordings and photos that can be used for identity theft, user manipulation or financial fraud. The development of these technologies represents a serious threat to the privacy of individuals, the protection of personal data and digital trust (Okolie, C., 2023). In the financial sector, deepfake content can be used to impersonate company executives, authorize financial transactions, or bypass biometric authentication systems.

In addition to digital identity abuse, artificial intelligence enables the automation of various forms of cybercrime. Modern AI tools can analyze the vulnerabilities of information systems, automatically generate phishing messages tailored to specific users, create malicious program code and optimize attacks on financial institutions. The misuse of artificial intelligence represents one of the key issues of national security, as it enables the execution of complex cyber operations with minimal human involvement (Montasari, R., 2022). Such attacks can have serious consequences for the stability of the financial system, especially if they target critical infrastructure or electronic payment systems.

An additional dimension of the problem is represented by the activities of extremist and terrorist organizations that use artificial intelligence to automate propaganda, manipulate information and plan attacks. The development of generative artificial intelligence significantly increases the possibilities of misuse of digital technologies for extremist and terrorist purposes. Although these activities are not directly related to financial fraud, they indicate a broader security context in which AI systems are developed and used (Đorić, M., et al., 2021). Financial institutions that have large amounts of sensitive data are potential targets of such activities, which is why security risk management must be an integral part of AI system development.

The fight against the criminal use of artificial intelligence cannot be based solely on the development of new fraud detection technologies. It is necessary to establish a comprehensive framework that includes international standards, regulatory regulations, exchange of information between institutions and continuous education of employees. It is especially important that standardization represents one of the most effective mechanisms for preventing the misuse of AI systems, as it enables the definition of unique requirements for their safe implementation (Ayyaswamy, K., et al., 2025).

The development of artificial intelligence has significant implications in the field of defense and national security. Artificial intelligence represents one of the key technologies of future military systems, whereby the same technological solutions can be adapted for civilian purposes, including the financial sector (Nazil, A. R., 2025). This fact further confirms the need for international standardization and a harmonized regulatory approach to prevent uncontrolled or unethical use of intelligent systems.

Good practices for the development of secure digital products are of special importance for the security of the digital environment. The application of international standards, security guidelines and appropriate development procedures significantly contributes to reducing the risk of cyber attacks and limiting the possibility of conducting offensive cyber operations. These conclusions can be directly applied to artificial intelligence systems used in the financial sector, since the security of AI models depends not only on the quality of the algorithms but also on the security of the entire digital environment in which they are implemented (Radunović, V., 2026).

The aforementioned challenges confirm that standardization has a much broader role than defining technical requirements. It represents the basis for the development of reliable, safe and ethically acceptable AI

systems that can respond to modern threats in the digital financial environment.

8. DISCUSSION

A review of the analyzed literature shows that the application of artificial intelligence in the financial sector is developing much faster than the regulatory mechanisms that should ensure its safe use. Although machine learning and deep learning algorithms achieve exceptional results in the detection of financial fraud, their effectiveness is not the only criterion for evaluating the success of an AI system. Equally important are transparency, the ability to explain decisions, privacy protection, risk management and compliance with regulatory requirements.

Analysis of international standards shows that ISO/IEC 42001 represents a significant step forward in the establishment of an artificial intelligence management system. Combined with the ISO/IEC 23894 and ISO/IEC 38507 standards, organizations are enabled to develop a comprehensive framework that includes risk management, management accountability, data security and continuous improvement of AI systems.

On the other hand, the AI Act of the European Union introduces legally binding requirements for high-risk AI systems, including financial fraud detection systems. Thus, for the first time, a direct connection between international standardization and regulatory law is realized. It can be expected that the future development of international standards will be largely harmonized with the requirements of European regulations.

A special challenge is the fact that the same technology can be used both to improve security and to carry out criminal activities. This is precisely why the future development of AI systems must be based on the principles of "Security by Design", "Privacy by Design" and "Ethics by Design", which imply that security, ethical and regulatory requirements are integrated already at the system design stage.

Based on the conducted analysis, it can be concluded that the successful application of artificial intelligence in the financial sector will depend on the ability of organizations to simultaneously develop advanced algorithms, apply international standards and harmonize business with regulatory requirements.

The inclusion of controlling, internal auditing and Benford's law complements the two-level regulatory model. At the organizational level, controlling and internal auditing ensure continuous monitoring of AI system objectives, risks and compliance. On an analytical level, Benford's Law is a transparent statistical indicator that can help initially isolate unusual financial data. Their integration with machine learning methods contributes to a hybrid approach where automated detection remains verifiable, documented and under human supervision.

9. CONCLUSION

Artificial intelligence represents one of the most important technologies of the modern financial sector and plays a key role in improving the system for detecting and preventing financial fraud. The application of machine learning algorithms, deep learning and analysis of large data sets enables financial institutions to more effectively identify suspicious transactions, reduce operational costs and improve risk management.

However, the growing application of AI systems also brings new challenges related to transparency, accountability, data protection, ethical aspects and cyber security. Therefore, standardization is a key regulatory mechanism that establishes common rules for the development, implementation and management of artificial intelligence systems.

Analysis of international standards ISO/IEC 42001, ISO/IEC 23894, ISO/IEC 38507 and ISO/IEC 22989, together with the regulatory framework of the European Union and the NIST AI RMF guidelines, shows that the modern approach to regulating AI systems implies the integration of technical, organizational, legal and ethical requirements. Such an approach enables the development of reliable and transparent systems that simultaneously achieve a high level of efficiency and compliance with regulatory requirements.

Based on the conducted literature review, it can be concluded that standardization will not represent a limitation of the development of artificial intelligence, but one of the basic prerequisites for its sustainable and responsible application. In the coming period, further convergence of international standards and regulatory regulations can be expected, which will contribute to greater user confidence, increased security of financial systems and more effective suppression of financial fraud.

Additionally, linking standardization with controlling, internal auditing and statistical methods, such as Benford's law, enables the practical implementation of the requirements of responsible artificial intelligence. Controlling provides future direction and monitoring of results, internal audit an independent assessment of compliance and controls, while Benford analysis provides an understandable indicator of possible anomalies. Their joint application can improve the reliability, explainability and auditability of AI systems for financial fraud detection.

References

[1] Ayyaswamy, K., Kathirvel, N., & Vianny, M. M. (2025). *Strategies for combating criminal use and abuse of artificial intelligence*. In *Avoiding Ad Fraud and Supporting Brand Safety: Programmatic Advertising Solutions* (pp. 257–282). IGI Global Scientific Publishing.

[2] Blind, K., Jungmittag, A., & Mangelsdorf, A. (2011). *The economic benefits of standardization*. DIN German Institute for Standardization.

[3] Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.

[4] Đorić, I., Ranisavljević, M., & Kocka, Đ. (2024). *Benfordov zakon–plan ili planiranje i analiza*

planiranih veličina u budžetu odbrane. Trendovi u poslovanju, *1*(23), 39-47.

[5] Đorić, M., & Milošević, T. (2021). *Abuse of Artificial Intelligence for Extremist and Terrorist Purposes*. *Srpska politička misao / Serbian Political Thought*.

[6] Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers, 24*(5), 1709–1734.

[7] European Union. (2024). *Regulation (EU) 2024/... laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).*

[8] ISO/IEC 22989:2022. *Artificial intelligence – Concepts and terminology.*

[9] ISO/IEC 23894:2023. *Artificial intelligence – Guidance on risk management.*

[10] ISO/IEC 38507:2022. *Information technology – Governance implications of the use of artificial intelligence by organizations.*

[11] ISO/IEC 42001:2023. *Information technology – Artificial intelligence – Management system.*

[12] Montasari, R. (2022). Cyber threats and national security: The use and abuse of artificial intelligence. In *Handbook of Security Science* (pp. 679–700). Springer.

[13] Nazil, A. R. (2025). *AI at War: The Next Revolution for Military and Defense.*

[14] NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).*

[15] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection. *Decision Support Systems, 50*(3), 559–569.

[16] Okolie, C. (2023). Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies, 25*(2), 11.

[17] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review, 34*(1), 1–14.

[18] Radunović, V. (2026). *The Impact of Good Practices on the Security of Digital Products and the Limitation of Offensive Cyber Operations*. Univerzitet Singidunum.

[19] Zielke, T. (2020). Is artificial intelligence ready for standardization? In *European Conference on Software Process Improvement* (pp. 259–274). Springer.

INDEX OF AUTHORS

A

437 Adrijana Jovanović
114, 265, 908 Aleksa Aničić
733 Aleksa Grubić
2 Aleksandar Bengin
761, 773 Aleksandar Ćitić
679 Aleksandar Dimić
121, 138 Aleksandar Đurić
159 Aleksandar Kari
82 Aleksandar Knežević
605 Aleksandar Pluškoski
820 Aleksandar Rajković
820 Aleksandar Savić
880 Aleksandar Simonović
420 Aleksandar Stefanović
863, 902 Aleksandar Tanasković
132 Aleksandra Obradović
675 Aleksija Đurić
8, 14, 77 Aljoša Ljuboja
735 Ana Alil
562, 568 Ana Matović
679 Ana Petrović
152, 190 Ana Simonović
393 Anđa Denić
695 Anđela Simović
287 Anđela Ivković
431 Andrija Dević
89 Andrija Ostojić
109, 152 Andrija Veljović
101, 470 Anita Klikovac
740, 744 Anja Antanasković
460 Anja Perić
748 Anton Radojković

B

666 Bernd Friedrich
27, 33, 39 Biljana Ilić
247 Biljana Jovanović
767 Biljana Mihajlović
226 Blaža Stojanović
880 Boban Bogićević
358, 413 Boban Bondžulić
358, 413 Boban Pavlović
339, 345 Boban Sazdić-Jotić
573, 598 Boban Stojanović
59 Bogdan Bogdanović
213 Bogdan Živković
315 Bojan Šmidling
144, 185 Bojan Tanaskovski
863, 902 Boris Vakanjac
475 Borislav Đorđević
407, 494 Borko Đaković
444 Boško Božilović
645 Božana Gaćić
27, 33, 39 Božidar Simović
132 Branislav Nikolić
8, 14, 77 Branislav Ostojić
271 Branka Džunić
755 Branka Jordović
529 Branko Vujatović

C

401 Calin-Octavian Micloshina

D

283, 292 D.A. Anufrieva
283, 292 D.A. Saperov
114, 132, 908 Damir Jerković
308 Damir Krklješ
675 Damjan Klobčar
240, 277, 303 Danica Bajić
247 Danica Jelić
857 Danijela Kovač
247 Danijela Obradović
54 Danilo Vukosavljević
529 Darko Grubač
96, 254 Darko Janković
339, 371 Darko Mikanović
465, 482, 488 Darko Pijević
254, 633 Darko Vasiljević
465, 488 Dejan Ivković

89 Dejan Jevtić
666 Dejan Jovanović
857 Dejan Marsenić
670, 684 Denis Dinić
377 Dijana Damljanović
358, 413 Dimitrije Bujaković
517 Dimitrije Kolašinac
844 Dobrivoje Ćatić
271, 425 Dragan Knežević
494 Dragan Kondić
204 Dragan Lazić
675 Dragan Milčić
465, 488 Dragan Nikolić
121, 711 Dragan Stojadinović
173, 833 Dragan Trifković
345, 371 Dragana Bojić
247 Dragana Jelić
707 Dragana Vasić Anićijević
21 Dragoljub Tanović
815 Dragomir Glišić
233 Dragoslav Petrović
639, 786, 792 Dušan Arsić
247 Dušan Blagojević
833 Dušan Cincar
902 Dušan Polomčić
639 Dušan Popadić
666 Duško Kostić

Đ

591 Đorđe Đermanović
717 Đorđe Gjumišev
700 Đorđe Ivković

E

562 Edis Mekić
568 Elmedin Biberović
585 Emina Kričković
833 Enis Dauti
815 Esmail Ali Salem Ahmed

G

144, 185, 271 Goran Marjanović
333 Goran Marković
27, 33, 39 Goran Ocokoljić
431 Goran Vasilić
2 Goran Vorotović

H

226 Horia Abaitancei
863 Hristina Petrova

I

283, 292 I.N. Ulyanov
296 Igor Babić
605 Igor Ciganović
933 Igor Đorić
114 Igor Epler
502, 875 Igor Goršić
401 Igor Medenica
551, 573, 598 Igor Petrović
109 Igor Radisavljević
915 Igor Stamenković
908 Ilija Ljujić
377 Ilija Nenadić
444, 612 Ilija Popadić
482, 488 Ina Masnikosa
727 Irina Ulanova
27, 33, 39 Isidora Stanković
277 Ivan Dimitrijević
2 Ivan Kostić
201 Ivan Marić
352 Ivan Pokrajac
551 Ivan Tot
444, 612 Ivan Vulić
327, 691 Ivana Perović
377, 393 Ivana Petković
431 Ivana Vasović Maksimović

J

695 Jelena Bajat
507 Jelena Bajić
731 Jelena Lović
670 Jelena Dikić
327,691 Jelena Georgijević
780, 798 Jelena Gržetić
277, 287 Jelena Marinković
96, 271, 727 Jelena Mojsilović
744 Jelena Nikolić
810 Jelena Orelj
185, 271, 425 Jelena Petković-Cvetković
437 Jelena Popović-Božović
54 Jelena Šobot
240 Jelena Šultans
21, 65 Jelena Svorcan

401, 605 Jelena Vasiljević
382, 431 Jelena Vidaković
517 Jordan Atanasijević
620, 626 Jose F. M. L. Mariano
512 Jovan Radivojević
679 Jovana Antić
727, 773, 810 Jovana Mandić
213 Jovanka Lukić
717 Jovanka Pejić

K

639 Katarina Dragutinović
661 Katarina Kokanov Stanković
425 Katarina Mišković
96, 240, 633 Katarina Nestorović
839 Konstantin Komatina
517, 556 Kristina Živanović
482 Ksenija Bokić

L

254 Lana Milosavljevic
892 Lazar Bjelica
908 Lazar Blagojević
512 Lazar Velinov
633 Luka Marjanović

LJ

755 Ljubica Radović
254 Ljubiša Tomić

M

179, 523 Maja Rosić
761, 804 Maja Vitorović-Todorović
915 Marija Baltić
857 Marija Borović
885 Marija Janković
562, 568 Marija Matović
165 Marija Milovanović
27, 33, 39 Marija Samardžić
420 Marijana Stojanović
101, 470 Marina Dabetić
717 Marina Dojčinović
767, 804 Marina Ilić
48 Marina Ostojić
96, 254, 633 Marina Simović – Pavlović
857 Marinko Aleksić
77 Mario Silvagni
420 Marko Bojanić
773 Marko Bojin
700, 786, 792 Marko Delić
652, 827 Marko Milovanović
551, 573, 598 Marko Premović
48 Marko Prošić
271, 727 Marko Radisavljević
591 Marko Simić
152, 159, 190 Marko Simonović
254 Martin Ocsenhofer
65 Mašan Vučinić
591 Matija Milić
645 Mihael Bučko
220 Mihailo Bojković
226 Mihailo Đuković
296 Mihailo Erčević
579 Mihailo Knežević
308, 382 Mihailo Lazarević
213 Mihajlo Vasiljević
226 Milan Bukvić
556 Milan Gnjatović
902 Milan Kresojević
820 Milan Mihajlović
740 Milan Milivojević
425 Milan Popović
72 Milan Rakić
545 Milan Ristić
152, 159, 190 Milan Simonović
670, 684 Milan Tanić
494 Milena Grdović
327, 691 Milena Pijović Radovanović
413 Milenko Andrić
14 Milenko Trifković
679 Mileta Ristivojević
885 Milica Ćurčić
21, 915 Milica Ivanović
645, 731 Milica Košević
425 Milica Marjanović
365, 717, 722 Milica Vlahović
444, 612 Miljan Vučetić
755 Miljana Popović
114, 265 Miljko Pavlović
551, 556 Milomir Jovićević
815 Miloš Čenić
488 Miloš Jevtić
265, 908 Miloš Jončić
204, 401, 605, 871 Miloš Jovanović
89 Miloš Marković

488 Miloš Otović
875 Miloš Pavić
639 Miloš Pešić
173, 179, 523 Miloš Sedak
773, 850 Miloš Stepanović
928 Miloš Stojanović
44 Miloš Ugrenović
327, 691 Mina Seović
144 Miodrag Ćurčin
265 Miodrag Lisov
675 Miodrag Milčić
77, 8 Miodrag Milenković-Babić
321, 352 Miodrag Vračar
240, 259, 277 Mirjana Krstović
259 Mirjana Milekić
666 Mirko Stamatović
185 Miroslav Gluvačević
121 Miroslav Jovanović
928 Miroslav Mitrović
880 Miroslav Pajčin
871 Miša Markuš
666 Mitar Perušić
684 Mladen Nikolić
722 Mladen Pejić
277, 287 Mladen Timotijević
127, 551, 573 Mladen Veinović
388 Mladen Živković
138, 333 Momir Drakulić
358 Momir Stanković
303 Muhamed Sućeska

N

254 Nada Citakovic
259 Nada Marinković
700 Nada Ratković
717, 722 Nataša Đorđević
748 Nataša Pajić
675 Nataša Zdravković
928 Nebojša Nikolić
114, 132, 908 Nebojša Hristov
420 Nebojša Jovičić
810 Nebojša Mitrović
928 Nebojša Nikolić
82 Nebojša Radojković
871 Neda Spasojević
670 Nemanja Jovanović
308 Nemanja Tanasković
857 Nemanja Vujović
197 Nenad Kovačević
815 Nenad Radović
494 Nenad Stefanović
529 Nenad Stojanović
365, 722 Nenad Vušović
413 Nikola Bukvić
407 Nikola Danilović
786, 792 Nikola Kastratović
880 Nikola Kilibarda
786, 792 Nikola Komatina
127 Nikola Latinović
488 Nikola Luković
132 Nikola Mačar
857 Nikola Nikolić
652, 827 Nikola Prodanović
152, 159, 165 Nikola Radovanović
535, 540, 545 Nikola Vukotić
740 Nikola Vuković
54 Nikola Zdravković
382, 431 Nikola Živković
475 Nikola Zogović
450 Nina Mitričević
921 Novak Roganović

O

27, 33, 39 Ognjen Babić
8, 880 Ognjen Peković
277 Ognjen Rajković
460 Olga Jakšić
780, 798 Olga Maraš
2 Olivera Kostić

P

502, 579, 810 Pavel Crnomarković
898 Pavle Ljubojević
190 Pera Srbul
327, 691 Petar Laušević
875 Petar Mišljen
798 Petar Vasiljević
885 Predrag Božović
173 Predrag Dobratić
247 Predrag Elek

R

450, 455 Rade Pavlović
321, 345, 371 Radiana Brusin
666 Radislav Filipović

579, 810 Radoslav Surla
645 Radovan Karkalić
612 Ranko Petrović

S

283 S.S. Kuzmin
101, 470 Sanela Veljković
670 Sanja Jevtić
529 Sanja Klajić
735 Sanja Martinović
152 Sanja Milinković
352 Sara Milanović
780, 798 Sara Pepić
201 Saša Antonović
585, 591, 921 Saša Bakrač
502, 579, 875 Saša Gundelj
226 Saša Milojević
287 Saša Savić
679 Saša Spasenić
711 Saša Veselinović
96 Saša Živković
475 Sava Stanišić
921 Siniša Drobnjak
827, 652 Siniša Gaćić
722 Slavica Mihajlović
535, 540 Slavimir Stošović
556 Slaviša Ilić
591 Slaviša Tatomirović
82 Slaviša Vlačić
185, 271 Slobodan Bosiljčić
850 Slobodan Malbašić
265 Slobodan Orašanin
127 Slobodan Pelemiš
226 Slobodan Savić
407 Slobodan Simić
871 Slobodan Subotić
327, 691 Snežana Brković
744 Snežana Grujić
121 Snežana Jovanović
915 Snežana Mališić
377, 393, 933 Snežana Zurovac
804 Sonja Bauk
722 Srđan Matijašević
488 Srđan Radovanović
921 Srđan Stojaković
144 Srđan Stojković
502 Srđan Trandafilović
892, 850 Srđan Živković
815 Srđa Perković
666 Srećko Stopić
121, 645 Sreten Perić
679 Sreten Spasenić
303 Stanisław Cudziło
679, 815 Stefan Dikić
517 Stefan Ivanović
165 Stefan Jovanović
327, 691 Stefan Mitrović
315, 488 Stefan Stojkov
711, 54 Stevan Jovičić
197 Stevan Mušicki
670, 684 Stevan Stupar
321 Stevo Vračar
259 Suzana Kovačević
209 Svetolik Spasenić

T

661 Tamara Lazarević-Pašti
48 Tamara Nikolić
661 Tamara Terzić
744 Tamara Vujatović-Velimirov
482 Tanja Lacmanović
898 Tatjana Lazović
767 Tatjana Marković
740 Tatjana Šoštarić
735 Tatjana Volkov-Husović
240, 259, 780 Teodora Stančić
303 Thomas M. Klapötke
585 Tin Lukić
675 Tomaž Vuherer
14, 72 Toni Ivanov
59 Tonko Mihovilović

V

283, 292 V.S. Abrukov
339, 371 Vasilija Joksimović
661 Vedran Milanković
863 Veljko Marinović
744 Veljko Savić
727 Vera Joksimović
96, 727, 780 Veselin Živanović
773 Vesna Lukić
700, 786, 792 Vesna Mandić
240, 259 Vesna Petrović
863, 902 Vesna Ristić Vakanjac
197 Vinko Žnidaršič
620, 626 Violeta Nikolić

767 Vlada Petrovic
333 Vlada Sokolović
707, 661 Vladan Anićijević
585 Vladan Tadić
292 Vladica Bozic
740 Vladimir Adamović
844 Vladimir Ćatić
666 Vladimir Damjanović
220 Vladimir Dunić
220, 639 Vladimir Milovanović
744 Vladimir Topalović
109, 190 Vuk Garčević
173 Vuk Milanović
675 Vukašin Pavlović
767 Vukica Grković

Ž

204 Žaklina Čutović
898 Žarko Mišković
700 Živana Jovanović Pešić
773 Živorad Jevtić

Z

444, 612 Zoran Banjac
585 Zoran Kričković
437 Zoran Krstić
2 Zorana Dančuo
740, 744 Zorica Lopičić
465, 488 Zvonko Radosavljević

SPECIAL THANKS TO

IRITEL BEOGRAD