Technical University of Košice

Faculty of Electrical Engineering and Informatics

SCYR

26th Scientific Conference of Young Researchers
Proceedings from Conference

ISBN 978-80-553-5051-6

2026

Sponsors & Organizers

SCYR 2026: 26th Scientific Conference of Young Researchers
Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 264 pages, Available Online

Editors: Assoc. Prof. Ing. Karol Kyslan, PhD.
Assoc. Prof. Ing. Emília Pietriková, PhD.
Ing. Lukáš Pancurák, PhD.

ISBN 978-80-553-5051-6

Scientific Committee

General chair: Prof. Ing. Liberios Vokorokos, PhD.

Editorial board chairman: Assoc. Prof. Ing. Karol Kyslan, PhD.

Committee Members:

Ing. Martin Čertický, PhD., Caterpillar Slovakia
Ing. Tibor Gujdan, FPT Slovakia
Ing. Juraj Bujňák, PhD., Siemens Healthineers
Ing. Alexander Revák, Magna Electronics Slovakia
Ing. Ján Šepeľa, BSH Drives and Pumps
Prof. RNDr. Vladimír Lisý, DrSc.
Prof. Ing. Alena Pietriková, CSc.
Assoc. Prof. František Babič, PhD.
Assoc. Prof. Peter Feciľak, PhD.
Assoc. Prof. Stanislav Ondáš, PhD.
Assoc. Prof. Peter Papcun, PhD.
Assoc. Prof. Norbert Ádam, PhD.
Assoc. Prof. Martin Sarnovský, PhD.
Assoc. Prof. Emíla Pietriková, PhD.
Assoc. Prof. Matúš Pleva, PhD.
Assoc. Prof. Ing. Marek Pástor, PhD.

Organizing Committee

Members:

Assoc. Prof. Ing. Karol Kyslan PhD.
Assoc. Prof. Ing. Emília Pietriková, PhD.
Ing. Ivana Olšiakova
Ing. Lukáš Pancurák, PhD.
Ing. Peter Šofranko
Ing. Juraj Kováč
Ing. Daniel Šolc
Ing. Peter Nemergut
Ing. Michaela Leľová
Ing. Samuel Onufer
Ing. Richard Král
Ing. Matej Vaľko

Contact address:
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9
040 01 Košice
Slovak Republic

Proceedings Reviewers

František Babič
Norbert Ádám
Ján Bačík
Michaela Bačíková
Peter Bednár
Ľubomír Beňa
Matej Bereš
Štefan Berežný
Anna Biceková
Peter Bober
Samuel Bucko
Gabriel Bugár
Peter Butka
Zsolt Čonka
Miloš Drutarovský
Peter Duranka
Frantisek Durovsky
Jaroslav Džmura
Peter Feciľak
Pavol Fedor
Želmíra Ferková
Peter Girovský
Peter Gnip
Filip Gurbáľ
Milan Guzan
Renát Haluška
Máté Hireš
Daniel Hládek
Sergej Chodarev
Eva Chovancova
Patrik Jurík
Ondrej Kainz
Erik Kajáti
Ján Kaňuch
Jozef Király
Mária Kladivová
Štefan Korečko
Ondrej Kováč
Jozef Kravčák
Karol Kyslan
Milan Lacko
Tomáš Lenger
Vladimír Lisý
Ľubomír Livovský
Oliver Lohaj
Peter Lukács
Branislav Madoš
Ján Magyar
Marian Mach
Kristína Machová
Dušan Medveď
Miroslav Michalko
Monika Molnárová
Stanislav Ondáš
Jozef Onufer
Ľuboš Ovseník
Lukáš Pancurák
Ján Paralič
Marek Pástor
Marek Pavlík
Daniela Perdukova
Ján Perháč
Alena Pietriková
Emília Pietriková
Matúš Pleva
Marek Ružička
Martin Sarnovsky
Peter Sinčák
Branislav Sobota
Zuzana Sokolová
Ján Staš
William Steingartner
Matúš Sulír
Csaba Szabó
Martina Szabóová
Slavomír Šimoňák
Eugen Šlapak
Viktor Šlapák
Róbert Štefko
Martin Tomášek
Ľubomír Urblík
Ján Vaščák
Igor Vehec
Marcel Vološin
Peter Vrábel
Ján Zbojovský
Iveta Zolotova
Jaroslava Žilková

Foreword

Dear Colleagues,

the Scientific Conference of Young Researchers (SCYR) is a scientific event focused on exchange of information among young researchers from Faculty of Electrical Engineering and Informatics at the Technical University of Košice (FEEI TUKE). It is a series of annual events that was founded in 2000. Since 2000, the conference has been hosted by FEEI TUKE with rising technical level. The 26^{th} Scientific Conference of Young Researchers (SCYR 2026) was held on April 17, 2026 at the University Conference Centre, Technical University of Košice. The mission of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities of young researchers at the Faculty of Electrical Engineering and Informatics, has been achieved. In total, 70 participants presented their papers during the conference.

The Faculty of Electrical Engineering and Informatics has a long-standing tradition of engaging students in practical professional activities, where they are encouraged to apply their theoretical knowledge in real-world settings. The SCYR provides doctoral and graduating students with an opportunity to develop and refine their scientific communication and presentation skills. At the same time, its primary mission remains to serve as a platform for the exchange of ideas and research outcomes among young scientists from different departments and communities. The conference brings together experimental and theoretical contributions across a broad spectrum of scientific disciplines, including computer science, artificial intelligence, electrical engineering and physical engineering of advanced materials.

Traditionally, contributions was divided in 3 categories:

- Electrical Engineering and Advanced Materials,
- Artificial Intelligence,
- Computer Science and Related Topics,

This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the faculty and faculty partners and sponsors. This scientific network addressed various research and educational challenges while promoting communication doctoral students, young researchers, and professors. The conference also provided an opportunity to initiate future scientific cooperation.

The results presented in the papers demonstrated that the research activities conducted out by young scientists make a significant contribution to advancing the scientific, educational and technologic objectives of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice.

We would like to express my sincere gratitude to all participants for their valuable contributions to these proceedings. We hope that the conference continues to serve as a platform for ongoing dialogue and collaboration among young scientists.

It is our great pleasure and honour to thank our sponsors, faculty partners, colleagues, committee members, and all those whose ideas, discussions, dedication, and hard work contributed to the successful organization of this event. We also extend our appreciation to the session chairs for their professionalism, cooperation, and commitment throughout the conference.

Finally, we would like to thank all attendees for the inspiring discussions, positive atmosphere, and for making this event a truly rewarding experience.

Liberios VOKOROKOS
Dean of FEEI TUKE

April 17, 2026, Košice

Contents

Erik Podracký
Optimization of mild-hybrid drive: An overview 12

Tomáš Danko
Efficient Hardware Acceleration of Deep Neural Networks: A Review 17

Ladislav Hric
Augmentation method for robust UAV-based military object detection 21

František Margita
Weather-Based Spatial Dynamic Line Rating of Overhead Transmission Lines Using a CIGRE 601 Thermal Model 24

Jozef Badár
Artificial Intelligence for Routing in 6G Mobile Multi-Hop Networks: A Progress Report and Comparative Study 27

Renáta Rusnáková
Quantum Annealing in Traffic Management & Quantum-ML Synergy 30

Dušan Čatloch
Malware Threat Detection Using Machine Learning Methods 33

Eva Kupcova
Privacy-Preserving Biometric Authentication Using Advanced Cryptographic Techniques 36

Jerguš Ševec
Reconstructive Approach for Wearable Blood Pressure Monitoring 39

Daniel Šolc
Scene Understanding and Corner-Case Handling in Autonomous Driving: A Review 43

Nikola Tomaščíková
Global Optimization and Regularity in Discrete Dynamic Systems over Extremal Algebras 47

Kristián Eliáš
Research on the possibilities of increasing the integration of renewable energy sources into the electric power system 50

Michal Stromko
Autonomous Knowledge Synthesis: Multi-Agent Methodology for Systematic Scientific Literature Analysis 54

Vladimir Szomosi
Impact of V2G on Smart Grid 58

Július Šimčák
Resilient Energy Management in Smart Buildings: A V2G-Integrated DC Nanogrid Approach ... 62

Viliam Balara
Detection of AI-Generated Essays 65

Martin Eliáš
Large Reversible Strain in Ni-Fe-Ga Heusler Microwires 68

Lukáš Tomaščík
Emerging Architectures in Modeling and Simulation of Complex Physical Systems using Neural Networks 71

Zuzana Liščinská
Hybrid FSO/RF Link Management Using RSSI Prediction and Reinforcement Learning 74

Juraj Kováč
Current State and Innovative Concepts in Technological Support for Diagnosis and Therapy of Speech and Hearing Disorders 77

Peter Demeter
Multimodality of Transformer-Based Models for Toxicity Detection in Online Games and Social Networks 81

Svitlana Syzranova
A Survey of Anomaly Detection Techniques for Distributed IT Systems 85

Chris Uličný
Joint Split Computing and Early Exiting for Adaptive Transformer Inference in Vehicular Edge Environments 89

Simona Chovancová
Challenges and Future Directions for Explainable AI in Medical Imaging 93

Sylvia Maťašová
Research and Development of Artificial Intelligence for Indexing and Structuring Medical Data . 97

Matej Vaľko
Temperature and Pressure Effects on Breakdown in Nanofluids 100

Jaroslav Marko
Multi-Metric Patch Prioritization for Small Networks Using Attack-Graph Context 104

Timotej Sobota
Advanced XR technologies, their implementation environments, and applications 107

Ivana Nováková
Bayesian Multi Sensor AMP-vSLAM for Robust Industrial AR 111

Matúš Dopiriak
Adaptive Sensor Fusion for Semantic Corner Case Detection Using Agentic AI 114

Peter Nemergut
Sb-Based Solders for Large-Area Joints in Power Electronics: Reflow Profile Optimization 117

Lenka Bubeňková
Contextual Gamified Cybersecurity Education 120

Filip Mahdal
A Review of Ultra-Wideband Antennas on Paper-Based Substrates 123

Monika Kohútová
State Representation in Domain-Specific Languages: A Review of Semantic Models 127

Martin Nguyen
Automated Detection of Aurora Borealis in Mid-Latitudes Using Public Camera Networks 131

Miroslav Jusko
Analog Frontend for Measuring Bioimpedance 134

Nazarii Herasymchuk
Using visualizations to increase teaching effectiveness in the field of algorithms and data structures 137

Miroslav Murin
A Multimodal Deep Learning Framework for Detecting Hoaxes and Fake Images in the Digital Environment 141

Samuel Fodor
Deep learning methods for natural language processing 144

Tomáš Buček
Quantifying Student Contributions in Collaborative Programming Projects Through Advanced Git-Based Metrics 148

Tomáš Basarik
Decoupling Control of Five-Port MMAB Converter Based on the Newton-Raphson Method 151

Daniel Gordan
Pushing PSFB technology to 500 kHz: What is gained and what is lost? 154

Michaela Leľová
Multi-Agent Reinforcement Learning in Energy Markets: State and Outlook 157

Roland Molnár
Bidirectional DC/DC Converters: An Overview 161

Marek Horváth
Decomposing the Programmer Attribution Problem in Educational Source Code 165

Peter Šofranko
Sensorless PMSM Control via High-Frequency Signal Injection: Principles and Methods 168

Samuel Onufer
Switching Field and Magnetoelastic Properties of Amorphous Glass-Coated Microwires 172

František Kurimský
Analysis and Modeling of Electricity Market Dynamics under Increasing Renewable Integration 176

Alex Krempaský
Large Language Models in Energy Commodity Trading 179

Ardian Hyseni
Convolutional NN and Benford's Law in Electricity Network Anomaly Detection 183

Ľubomír Pallaj
Machine Learning Approaches for Partial Discharges Classification in High-Voltage Equipment 186

Branislav Setlák
Computer Vision Foundation Models in Medical Imaging 190

Miroslav Imrich
DRL-Based RIS Optimization for Smart Factories Leveraging Radiance Field Digital Twins ... 194

Peter Poprik

Towards Procedural Generation of Degraded Content Using Hierarchical Data Structures 197

Sebastian Hresko
Dysarthric Speech Detection in Parkinson's Disease Patients Using Foundation Models and Domain Generalization Techniques 200

Antónia Kováčová
Sparse Wars – Return of Donoho 205

Richard Král
Towards Affordable Wearable 3D LiDAR Navigation Assistance: Scanner Design and Calibration 208

Matej Hric
Possibilities of using Digital Twin for flying shears 211

Dominika Líšková
Contribution to modelling and analysis of cyber-physical systems within the framework of the ALICE Detector Control System 214

Tadeáš Kmecik
Analysis, Identification, Modeling and AI-Based Compensation of Cogging Torque 217

Kristián Glajc
Progress in experimental investigation of discharges in nanofunctionalized transformer oi 220

Kristian Sopkovic
Morphological and Syntactic Analysis of Synthetic Questions Generated by LLM Models for Slovak Wikipedia 223

Erika Abigail Katonová
Dynamic Routing QoS-Guaranteed Software Defined Network Architecture 226

Kamil Ševc
The Impact of Photovoltaic Systems on Power System Management 231

Miroslav Komorník
Ing. 234

Tomáš Kormaník
Analysis of IoT Landscape in Education 238

Samuel Novotný
Transparent Intensional Logic from a Programming Language Perspective 241

Marek Bobček
Shadow Layer Protection: A Software-Defined Superstructure for Resilient Smart Grid Infrastructure 244

Norbert Rešovský
Solid-State Transformers in AC-Fed Locomotives: Overview and Future Developments 248

Tomáš Kmec
Evaluation of Adhesion Strength of Atmospheric Plasma Sprayed Coatings 253

Tomáš Kmec
Integrated Adhesion Enhancement of Atmospheric Plasma Sprayed Alumina Coatings Through Substrate Surface Geometry Modification 256

Tomáš Kmec
Necessity of Sealing for Atmospheric Plasma Sprayed Alumina Insulation Coatings Against Humidity 259

Author's Index 262

Optimization of mild-hybrid drive: An overview

[1]*Erik PODRACKÝ (1st year),*
Supervisor: [2]*František ĎUROVSKÝ*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Košice, Slovak Republic

[1]erik.podracky@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*—This paper presents a comprehensive review of the current state-of-the-art in mild-hybrid electric vehicle (MHEV) powertrain technologies and energy management strategies (EMS). As global regulations mandate drastic reductions in greenhouse gas emissions, 48 V electrification has emerged as a cost-effective transitional solution. This review categorizes hybrid architectures—series, parallel, and power-split—with a specific focus on the technical implications of P0 to P5 topologies. Key hardware components, including lithium-ion energy storage, bidirectional DC/DC converters, and belt starter generators (BSG), are analyzed in the context of their contribution to fuel economy. Furthermore, the integration of supervisory control approaches, from heuristic methods to optimization-based strategies like model predictive control (MPC) and equivalent consumption minimization strategies (ECMS), can reduce CO_2 emissions by up to 15 % in standard driving cycles while maintaining system complexity and manufacturing costs at viable levels.**

***Keywords*—48 V architecture, energy management strategy, mild-hybrid electric vehicle, powertrain topologies.**

I. Introduction

The global automotive sector is currently undergoing a transformation driven by the need for decarbonization and climate change mitigation [1]. Increasingly stringent regulations, such as the United Kingdom's net-zero target by 2050 and the projected dominance of electrified vehicles by 2040, are accelerating this shift [2]. While battery electric vehicles (BEVs) offer zero tailpipe emissions, their adoption is limited by charging infrastructure and battery energy density. Therefore, hybrid powertrains represent an effective transitional solution, balancing emissions reduction, cost, and driving range [3], [4].

Mild-hybrid electric vehicles (MHEVs), particularly 48 V systems with belt starter generator (BSG - in P0 topology), are among the most viable solutions for reducing fuel consumption and greenhouse gas emissions [5], [6], [7]. They enable fuel savings of approximately 10-15 % in standard cycles like the New European Driving Cycle (NEDC) through functionalities such as stop-start, torque assist, and others [6]. Additionally, electrification improves internal combustion engine (ICE) efficiency by shifting operation toward optimal regions of brake-specific fuel consumption (BSFC) [8].

A key challenge lies in the optimization of energy management strategies (EMS), which must coordinate power flows in real time to maximize efficiency [4], [9]. Advanced control approaches, including model predictive control (MPC) and equivalent consumption minimization strategies (ECMS), are essential for adapting to dynamic driving conditions [8], [10], [11]. Effective integration of hardware and control systems is crucial for meeting future emission targets while maintaining vehicle performance and cost-effectiveness [8].

II. Powertrain architecture of mild-hybrid vehicles

The classification of electrified powertrains depends on the energy flow topology and the integration of primary and secondary power sources [4], [9], [12]. Hybrid electric vehicles (HEVs) are typically categorized as series, parallel, or series-parallel (power-split) based on ICE and electric motor (EM) arrangement [13], [14].

A. Series hybrid

In a series hybrid configuration, the ICE is not mechanically connected to the drive wheels [13]-[15]. This configuration enables the engine to operate at an optimal efficiency point independently of vehicle speed [12], [14], [15].

However, multiple energy conversions result in increased losses and necessitate larger electrical components, making this architecture unsuitable for cost-sensitive mild-hybrid applications [13], [14], [16], [17], [18].

B. Parallel hybrid

The parallel hybrid architecture is the most prevalent configuration for MHEVs [16]. In this layout, both the ICE and the EM are mechanically coupled to the same drive shaft or transmission, allowing them to deliver propulsion power individually or simultaneously [9], [13]-[15]. Because the motor is connected directly to the drivetrain, the inefficiencies of dual energy conversion found in series hybrids are avoided, resulting in higher efficiency during high-speed highway cruising [12], [13]. Mild parallel hybrids typically utilize a low-power EM (often 5–20 kW) to provide all operating modes that MHEV can offer without the capacity for sustained pure electric propulsion [14], [16], [19].

C. Power-split hybrid

Power-split (or series-parallel) hybrids combine the advantages and complexities of both series and parallel systems [13], [15], [20]. This architecture utilizes a power-split device (PSD), most commonly a planetary gear set, to decouple the engine speed from the vehicle's road speed [9], [14], [15], [21]. One EM acts primarily as a generator to regulate the engine's operating point, while another serves as a traction motor [14], [15].

This configuration enables the vehicle to operate as a series hybrid at low speeds and as a parallel hybrid at higher speeds, effectively functioning as an electronic continuously variable transmission (e-CVT) [13], [15] but increases system complexity and cost, making it more suitable for full hybrid applications than for mild-hybrids [13], [20].

III. Categorization of Parallel Mild-Hybrid Vehicles

MHEVs are defined by a degree of hybridization (DoH) in which propulsion is primarily provided by the ICE and supported by a 5–20 kW EM [4], [6], [9], [20]. Unlike full hybrids, the EM in a mild configuration generally lacks the capacity for pure electric propulsion, focusing instead on optimization of the ICE through torque assistance and energy recovery [1], [4]. The 48 V architecture has become the industry standard due to its balance between cost and performance [5]. Figure 1 illustrates the DoH of MHEVs.

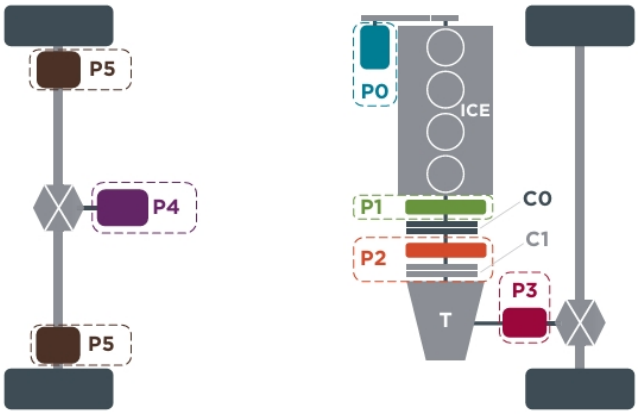


Fig. 1. Architectures P0 to P5 of parallel hybrid powertrains [22].

A. P0 architecture

The P0 topology is characterized by a belt-driven starter generator (BSG) integrated into the front-end accessory drive (FEAD) of the ICE [6], [23]. This arrangement replaces the conventional 12 V alternator with a more powerful 48 V EM [5], [9].

The P0 configuration represents the most prevalent market solution due to its minimal impact on existing vehicle platforms, requiring only enlarged pulleys and modified belt-tensioning systems to handle bidirectional torque flows. Its primary functions include enhanced stop–start, engine-off coasting, and limited kinetic energy recuperation [7], [24].

B. P1 architecture

In the P1 configuration, the EM—often an integrated starter generator (ISG)—is mounted directly onto the engine crankshaft, typically replacing the flywheel. This topology eliminates belt-related losses, offering higher efficiency in energy recovery and more robust torque assist compared to P0.

However, because the EM is rigidly coupled to the crankshaft, it cannot be decoupled from the engine's mechanical drag during recuperation, which limits its overall energy efficiency potential [9].

C. P2 architecture

The P2 architecture positions the EM between the ICE and the transmission input. A key technical advantage is the inclusion of a side-attached or coaxial clutch between the engine and the EM, allowing for complete decoupling of the ICE [9], [18]. This enables functions such as sailing (engine-off coasting) and improved regenerative braking, as the EM can harvest kinetic energy without the parasitic losses of the engine's internal friction [16].

D. P3 architecture

The P3 topology places the EM at the transmission output shaft, connected via gear mesh or a chain drive [9], [18]. This configuration provides the highest recuperation potential among pre-axle topologies because the electrical energy path is separated from the transmission's gear-shifting losses. The EM rotates at a multiple of the wheel speed, providing instantaneous torque response at the drivetrain level [9].

E. P4 architecture

The P4 architecture utilizes an electric axle drive, where the EM is mounted on the rear axle (or the axle not driven by the ICE). This layout effectively provides All-Wheel Drive (AWD) capabilities and allows for sophisticated torque vectoring to enhance vehicle stability and dynamics [6], [18].

P4 is considered highly effective for heavy-duty mild hybridization, as it requires no modifications to the primary front-end powertrain [18].

F. P5 architecture

The final, less commonly used topology is P5, which focuses on integrating motors directly into wheel hubs. [15], [20]. Each wheel is driven by an individual EM, offering the ultimate degree of freedom for torque control and packaging. While this eliminates the need for differentials and driveshafts, it presents significant engineering challenges regarding increased unsprung mass and environmental sealing of the motor units within the wheel assembly [15]. A comparison of individual MHEV architectures (P0–P5) is summarized in Tab. I.

IV. Components of Mild-Hybrid Vehicles

The integration of mild-hybrid technology requires high-performance components capable of bidirectional power flow while maintaining a balance between system cost and fuel economy benefits [24], [25]. The transition to 48 V architectures has standardized the primary hardware requirements, which include advanced energy storage, reversible EM, and sophisticated power electronics [6], [9].

A. Battery

The battery system in a mild-hybrid vehicle is optimized for high power density rather than high energy capacity, as it must support rapid bursts of torque assistance and high-rate energy recovery during regenerative braking. While early hybrid models utilized Nickel-Metal Hydride (NiMH) technology [20], modern 48 V MHEVs have universally adopted Lithium-ion (Li-ion) chemistry due to its superior specific power and energy density [9], [26]. Typical MHEV battery packs have a relatively low energy capacity, often less than 1 kWh [21], which is sufficient to facilitate extended stop–start routines and kinetic energy harvesting without the weight penalty of full hybrid batteries [20].

B. Electric machine

The EM is the central element of the hybrid powertrain, operating both as a motor for torque assist and as a generator for energy recuperation [7], [9], [18]. In MHEVs, it is typically rated between 7–15 kW [7]. Permanent Magnet Synchronous

TABLE I
COMPARISON OF P0–P5 ARCHITECTURES

Topology	Configuration	Cost/Complexity	Recuperation potential	CO_2 reduction
P0	BSG (FEAD)	Lowest (baseline)	Low (40 %)	3.8 % - 6.6 %
P1	Crankshaft-mounted	Low/Medium (+15 %)	Moderate (Limited by ICE)	8.5 %
P2	Pre-transmission	High (+30 % - 32 %)	High (ICE decoupling)	11.9 % - 15.8 %
P3	Post-transmission	High (+34 %)	Very high (No T-losses)	15.3 %
P4	Electric axle	Highest (+53 %)	Superior (Direct axle)	15.5 % - 23.9 %
P5	In-wheel motors	Highest (Unsprung mass)	Highest (Theoretical)	N/A (Emerging tech)

Machines (PMSMs) are most commonly used due to their high efficiency and power density [26], [27], [28], while Induction Machines (IMs) offer robustness and eliminate rare-earth materials [15], [27]. Emerging research also explores Switched Reluctance Machines (SRMs) for their high-power density and low production cost, though challenges remain regarding acoustic noise and vibration [9], [12].

C. Inverter

The traction inverter is the power electronic interface that regulates the energy flow between the DC battery and the AC electric machine [7], [16], [17]. It performs DC/AC conversion during motoring and AC/DC rectification during recuperation using high-frequency switching devices such as IGBTs or MOSFETs [14]-[16]. Controlled by a microprocessor unit, it applies established methods like PWM or vector control to ensure precise torque regulation and high efficiency [9], [12], [15].

D. DC/DC converter

A bidirectional DC/DC converter links the 48 V network and the 12 V board net [6], [18], [24], [26]. It supplies conventional 12 V loads (e.g., lights, sensors, infotainment) and charges the auxiliary battery [9], [15], [16]. In the reverse direction, the converter can allow the 12 V battery to assist the 48 V system during extreme cold-cranking conditions, ensuring reliable engine starts when the main Li-ion pack's performance is limited by temperature [9].

E. Additional components

Beyond the primary drivetrain, the 48 V system enables the electrification of power-demanding auxiliary components that significantly impact emissions and performance:

1) Electrically Heated Catalyst (EHC): This component uses electrical energy to rapidly heat the exhaust catalyst to its "light-off" temperature during cold starts, drastically reducing tailpipe NOx and HC emissions [5], [11], [20].

2) Electric Supercharger (e-Charger) / E-Turbo: An electrically driven compressor provides instantaneous boost pressure, effectively eliminating turbo lag and enhancing low-end torque [5], [6], [29].

3) High-Power Auxiliaries: Components such as electric water pumps, air conditioning compressors, and cooling fans can be decoupled from the engine speed, allowing for demand-controlled operation and further reducing parasitic losses [6], [14], [18].

V. OPERATING MODES OF MILD-HYBRID VEHICLES

The operational versatility of MHEVs is derived from the intelligent coordination of the ICE and the EM across diverse driving scenarios. These modes, shown in Figure 2, are designed to minimize fuel consumption and emissions while enhancing the vehicle's dynamic response and passenger comfort [5], [17], [24].

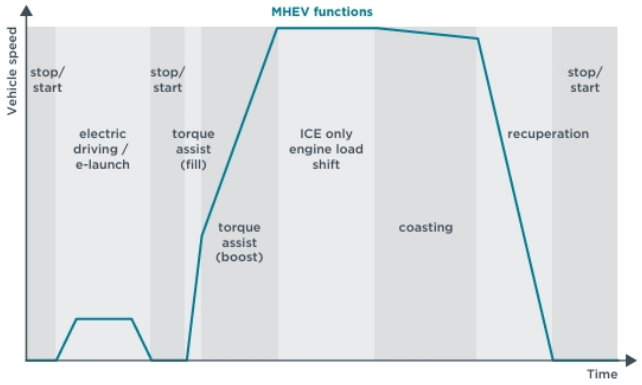


Fig. 2. Functionalities enabled by mild hybridization depending on vehicle dynamics [22].

A. Extended engine stop-start

An essential feature of MHEVs is enhanced stop–start functionality, which shuts down the ICE during vehicle standstill to eliminate idling losses [13], [20]. Extended functionality systems allow engine shutdown at low speeds (typically below 20 km/h) before a complete stop [5], [24]. The integration of BSG ensures fast, smooth, and low-vibration restarts compared to conventional starters [24], [29].

B. Recuperation of energy during deceleration and braking

During deceleration and braking, the EM operates as a generator, converting kinetic energy into electrical energy [5], [21]. This energy is stored in the 48 V battery instead of being dissipated as heat. Recuperation is particularly effective in urban driving, and in a P0 architecture, up to 40 % of deceleration energy can be recovered after conversion losses [18], [21], [24].

C. Increasing combustion-engine efficiency by shifting the load point

Load-point shifting, also referred to as load-leveling, optimizes the ICE's operation by moving its operating point into regions of lower BSFC (Figure 3). When the engine is operating at low loads with poor efficiency, the EM acts as a generator to increase the engine load, simultaneously

charging the battery [18], [29]. Conversely, when the engine faces unfavorable high loads, the EM provides torque to reduce the ICE's burden, thereby shifting the operating point back into the engine's "comfort zone" for better efficiency and lower emissions [5], [24], [29].

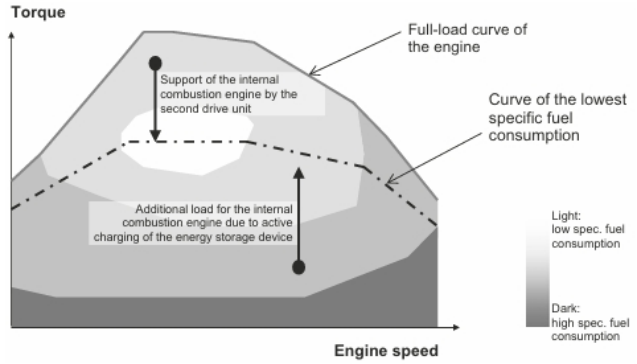


Fig. 3. The load-point shifting of the ICE [22].

D. Torque assist

The torque assist (or "boost") uses the EM to supplement ICE output during high-power demand, such as vehicle launch or rapid acceleration [24], [29], improving torque build-up and mitigating turbo lag in downsized engines [6], [24], [29]. The EM can also absorb part of the requested torque to reduce ICE peak loads known as "phlegmatization," lowering NO_x and soot emissions [5].

1) Torque fill: Due to its fast dynamic response and high low-speed torque, the EM compensates for the ICE's delayed torque buildup during transients, improving drivability and responsiveness (Figure 4). Due to mechanical and gas exchange inertia, the ICE is unable to respond instantaneously, particularly in the low-speed operating region.

Consequently, the application of EM torque fill enhances overall system responsiveness and improves vehicle drivability. Furthermore, torque fill can be combined with ICE downspeeding strategies to increase system efficiency while preserving adequate torque responsiveness [15].

2) Torque boost: Torque boost increases total system torque by combining ICE and EM output, especially at low speeds, supporting engine downsizing for better fuel economy and emissions. Operation is limited by battery and thermal constraints [15].

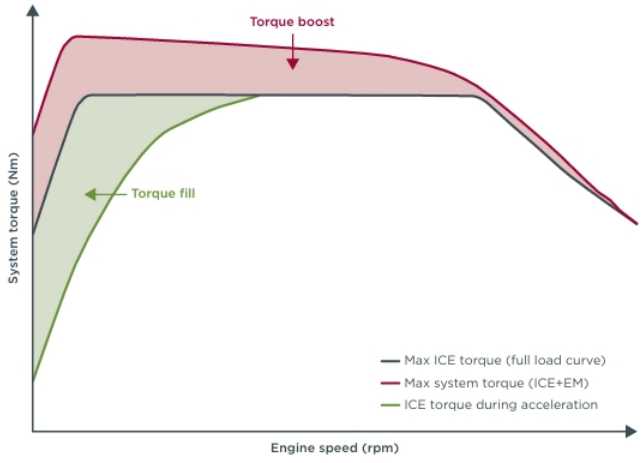


Fig. 4. Engine torque fill and torque boost functions [22].

VI. CO_2 Emissions Reduction and Fuel Saving Potential

The integration of 48 V mild-hybrid systems is driven by stringent global emission regulations, such as the European fleet-average CO_2 target of 95 g/km (2021), requiring further reductions of 15 % by 2025 and 37.5 % by 2030 [9], [23], [28]. Mild hybridization offers up to 70 % of full hybrid fuel economy benefits at roughly 30 % of the cost [6], with typical CO_2 and fuel savings of 10–15 % in standard cycles like NEDC, reaching 21 % in urban conditions [6], [16], [28].

The extended stop–start contributes 6.5 % savings compared to the 3–4 % for conventional 12 V systems [28], while in P0 topology regenerative braking recovers 40 % of deceleration energy [24]. Topology also affects the optimization ceiling: P0 is the most cost-effective, while P4 is the most fuel-efficient due to higher recuperation and EM decoupling (Tab. I).

Additional improvements arise from 48 V e-boosting and advanced EMS strategies (map-based ECMS), with load-shifting reducing NO_x by up to 22 %, soot by up to 36 %, and CO_2 (1.1 g/km) by preventing the ICE from operating in high-emission transient regions. [5], [6], [8].

VII. Energy Management Strategies

The EMS, implemented within the Vehicle Supervisory Controller (VSC), serves as the central intelligence governing the power flow between the ICE and the EM. Its objective is to meet driver torque demand while minimizing fuel consumption and emissions, maintaining charge-sustaining (CS) operation within a defined SoC window [8], [9], [13], [14], [20], [29].

A. Rule-based strategies

Rule-based (heuristic) strategies rely on deterministic or fuzzy logic to determine the operational mode based on real-time vehicle states such as speed, torque demand, and battery SoC [9], [21].

B. Optimization-based strategies

Optimization-based EMS aim to minimize a predefined cost function—typically total fuel mass or a combination of fuel and emissions—over a specific time horizon [11].

C. Model Predictive Control (MPC)

MPC solves a real-time optimization problem over a moving prediction horizon [10], [20]. By integrating Electronic Horizon data—including road gradients, curvature, and segment speed limits from navigation systems—MPC can proactively plan the SoC trajectory. This allows the powertrain to perform "load-leveling," where the ICE is shifted to its highest efficiency regions while the EM compensates for transient torque requirements. Although MPC requires significant computational resources, the development of high-efficiency solvers and virtual ECU technology has enabled its integration into mass-production hybrid control units [10].

D. Impact on Emission Management

Advanced EMS also incorporate strategies for emission mitigation, such as phlegmatization (torque-shaving), which reduces peak engine loads during transients to lower NO_x and soot generation. Furthermore, optimal control of the EHC can be integrated into the EMS to accelerate catalyst light-off during cold starts, significantly decreasing tailpipe emissions without excessive electrical energy waste [5], [11].

VIII. Conclusion

The systematic analysis of mild-hybrid powertrain technologies presented in this review confirms that 48 V systems represent a critical bridge between conventional ICE and high-voltage electrification. By utilizing low-voltage architectures that remain below the 60 V DC safety threshold (which is internationally recognized as the maximum safe touch voltage for DC) [5], [22], manufacturers can achieve significant improvements in fuel efficiency and pollutant emissions without the prohibitive costs associated with full hybridization. MHEVs also circumvent the stringent safety and regulatory requirements mandated for high-voltage vehicles (typically 400 V to 800 V), such as the need for specialized high-tension protection, orange-coded cabling, and galvanic isolation to prevent arcing or lethal electric shock. Furthermore, 48 V enables a fourfold increase in available power over traditional 12 V systems, allowing for the electrification of demanding auxiliary loads like EHC and e-Chargers [16].

Topology categorization shows that P0 configuration remains the most commercially prevalent due to its non-intrusive integration, while P2 and P4 offer better energy recovery by decoupling the EM from engine drag. MHEV functionalities—including extended stop–start, torque assist, and load-point shifting—reduce ICE peak loads, improve thermal efficiency, and lower transient emission spikes.

Furthermore, advances in EMS, from rule-based logic to predictive algorithms like MPC, together with Electronic Horizon data and real-time optimization, enable proactive SoC management and maximize benefits of downsized engines.

Future research should focus on addressing several open problems that are critical for the next generation of 48 V MHEVs:

1. Synergy with Euro 7 Standards: Investigating the optimal coordination between eCAT and ICE load-point shifting during the first 60 seconds of a cold start.

2. Predictive EMS and Electronic Horizon: Developing MPC algorithms that utilize Electronic Horizon data (road gradients and speed limits) for proactive SoC management.

3. Real-time Optimization: Implementing computationally efficient solvers to run optimization-based strategies on automotive microcontrollers in real-time.

Acknowledgment

This work was supported by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the project KEGA 059TUKE-4/2024 and by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00018.

References

[1] J. C. Castillo, A. F. Uribe, J. E. Tibaquirá, M. Giraldo, and M. Idárraga, "Hybrid electric vehicles as a strategy for reducing fuel consumption and emissions in latin america," *World Electric Vehicle Journal*, vol. 16, no. 2, 2025. [Online]. Available: https://www.mdpi.com/2032-6653/16/2/101

[2] G. Santos and C. A. Cornford, "Road transport electrification and motoring taxation in the uk," *Energy Reports*, vol. 11, pp. 4326–4337, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484723016086

[3] A. Kumar, D. K. Rajak, P. Kumar, and S. Rathee, Eds., *Smart Electric and Hybrid Vehicles: Design, Modeling, and Assessment by Industry 4.0 Approaches*. CRC Press, 2024, no. 1. [Online]. Available: https://doi.org/10.1201/9781003502470

[4] M. Iqbal, T. Wang, G. Li, D. Chen, and M. Al-Nehari, "A study of advanced efficient hybrid electric vehicles, electric propulsion and energy source," *Journal of Power and Energy Engineering*, vol. 10, no. 7, pp. 1–12, 2022.

[5] R. Romanato, F. Acquaviva, F. Duma, R. Fuso *et al.*, "48 v hybrid system technologies to develop the most efficient and cleanest diesel," SAE International, Tech. Rep. SAE Technical Paper 2018-37-0011, 2018. [Online]. Available: https://doi.org/10.4271/2018-37-0011

[6] B. Hu, C. Chen, Z. Zhan *et al.*, "Progress and recent trends in 48 v hybridisation and e-boosting technology on passenger vehicles – a review," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 232, no. 11, pp. 1543–1561, 2017.

[7] H. Malik, S. Mishra, Y. R. Sood, A. Iqbal, and T. S. Ustun, Eds., *Renewable Power for Sustainable Growth: Proceedings of ICRP 2023*, ser. Lecture Notes in Electrical Engineering. Springer Singapore, 2024, no. 1086. [Online]. Available: https://doi.org/10.1007/978-981-99-6749-0

[8] G. Cariccia, M. Pennazza, S. Brandin *et al.*, "Integrated design and control of p0 hybrid systems for co_2 reduction," *MTZ Worldwide*, vol. 83, pp. 30–37, 2022.

[9] I. Husain, *Electric and Hybrid Vehicles: Design Fundamentals*, 3rd ed. Taylor & Francis Group, LLC, 2021.

[10] J. Cho and H. Lee, "Mpc energy prediction control simulation of a hybrid electric vehicle," *International Journal of Automotive Technology*, vol. 26, pp. 607–619, 2025.

[11] A. Benaitier, F. Krainer, S. Jakubek, and C. Hametner, "Optimal control of aftertreatment electric heaters for mild hybrid vehicles during cold start," in *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2022, pp. 1–6.

[12] M. Ehsani, Y. Gao, S. Longo, and K. Ebrahimi, *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles*, 3rd ed. CRC Press, 2018. [Online]. Available: https://doi.org/10.1201/9780429504884

[13] Z. Namirian, "Comprehensive overview of hybrid vehicle drivetrains," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 8, aug 2020.

[14] L. Guzzella and A. Sciarretta, "Electric and hybrid-electric propulsion systems," in *Vehicle Propulsion Systems*. Berlin: Springer, 2007.

[15] C. Mi and M. A. Masrur, *Hybrid Electric Vehicles: Principles and Applications with Practical Perspectives*. Hoboken, NJ, USA: Wiley, 2017.

[16] T. Denton, *Electric and Hybrid Vehicles*. London, UK: Routledge, 2016.

[17] K. V. Singh, H. O. Bansal, and D. Singh, "A comprehensive review on hybrid electric vehicles: Architectures and components," *Journal of Modern Transportation*, vol. 27, pp. 77–107, 2019.

[18] M. Hilgers and W. Achenbach, *Alternative Powertrains and Extensions to the Conventional Powertrain*. Berlin: Springer Vieweg, 2021.

[19] A. B. Basavaradder, P. K. Dayanada, and K. N. Chethan, "Review on alternative propulsion in automotives – hybrid vehicles," *International Journal of Engineering & Technology*, 2018.

[20] G. Pistoia, Ed., *Electric and Hybrid Vehicles: Power Sources, Models, Sustainability, Infrastructure and the Market*. Oxford, UK: Elsevier, 2010.

[21] C. A. MacCarley, *Electric Vehicles*. Cham: Springer Nature Switzerland, 2025, pp. 391–572. [Online]. Available: https://doi.org/10.1007/978-3-031-85028-8_14

[22] J. Dornoff, J. German, A. Deo, and A. Dimaratos, "Mild-hybrid vehicles: A near term technology trend for co_2 emissions reduction," International Council on Clean Transportation (ICCT), Washington, DC, USA, White Paper, July 2022.

[23] K. Yoon, J. Hong, and J. Shim, "A study on front end auxiliary drive (fead) system of 48v mild hybrid engine," SAE International, Tech. Rep. SAE Technical Paper 2018-01-0414, 2018. [Online]. Available: https://doi.org/10.4271/2018-01-0414

[24] C. Helbing, K. Bennewitz, and A. Mann, "The 48-v mild hybrid drive system of the volkswagen golf 8," *MTZ Worldwide*, vol. 81, pp. 18–25, 2020.

[25] M. Lampalzer and M. Lechner, "48-v mild hybrid system for commercial vehicles," *ATZ Heavy Duty Worldwide*, vol. 13, pp. 22–26, 2020.

[26] F. Cerdas, *Integrated Computational Life Cycle Engineering for Traction Batteries*. Cham, Switzerland: Springer, 2022. [Online]. Available: https://doi.org/10.1007/978-3-030-82934-6

[27] C. H. T. Lee, W. Hua, T. Long, C. Jiang, and L. V. Iyer, "A critical review of emerging technologies for electric and hybrid vehicles," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 471–485, 2021.

[28] Z. Cao, A. Mahmoudi, S. Kahourzade, and W. L. Soong, "An overview of electric motors for electric vehicles," in *2021 31st Australasian Universities Power Engineering Conference (AUPEC)*, 2021, pp. 1–6.

[29] C. Kaup, S. Hoffmann, B. Unterberger *et al.*, "48-v diesel hybrid fun to drive at lowest co_2," *MTZ Worldwide*, vol. 77, pp. 18–23, 2016.

Efficient Hardware Acceleration of Deep Neural Networks: A Review

[1]*Tomáš DANKO (1st year),*
Supervisor: [2]*Norbert ÁDÁM*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]tomas.danko@student.tuke.sk, [2]norbert.adam@tuke.sk

***Abstract*—The increasing complexity of deep neural networks (DNNs) leads to higher computational requirements and increased power consumption. As a result, hardware acceleration is crucial to ensure that these models can be executed efficiently. This review examines existing hardware platforms, including CPUs, GPUs, FPGAs, CGRAs, and ASICs, as well as current software- and hardware-based techniques used to enhance computational efficiency. In addition, the paper identifies promising directions for future research, with a particular focus on reconfigurable and heterogeneous platforms.**

***Keywords*—Hardware acceleration, Deep neural networks, Graphics Processing Units, Field-Programmable Gate Arrays, Application-Specific Integrated Circuits.**

I. Introduction

Deep neural networks (DNNs) are computational models that enable deep learning (DL), an approach characterized by a layered structure that allows the progressive extraction of features from data [1]. Thanks to this functionality, DNNs form the backbone of many modern artificial intelligence (AI) applications and are widely used across domains such as automotive systems and healthcare. In numerous applications, DNNs are able to exceed human-level accuracy; however, this comes at the cost of high computational complexity [2].

This computational burden poses several challenges, including deployment in resource-constrained environments, slow and inefficient computation, and a high environmental footprint. These issues are becoming increasingly significant as DNNs continue to evolve rapidly and their computational and energy demands scale faster than improvements in general-purpose hardware. Hardware acceleration aims to address these challenges by employing specialized techniques that allow faster computation while optimizing resource utilization.

This review analyzes current research on hardware acceleration of DNNs, covering commonly used acceleration techniques and available hardware platforms, including general-purpose processors and specialized accelerators. The goal of this paper is to survey existing solutions and to identify promising directions for future research.

II. Literature Review and Analysis

The topic of hardware acceleration can be divided into two primary subareas: where the computation occurs and how it is achieved. The first subarea includes a range of technologies such as Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs), and Application-Specific Integrated Circuits (ASICs) [3]. Coarse-grained reconfigurable arrays (CGRAs), while less commonly used, also remain a relevant option. The latter subarea focuses on software- and hardware-based techniques that aim to enhance overall computational efficiency.

A. Hardware acceleration platforms

The most commonly available computing platform is the Central Processing Unit (CPU). It is a general-purpose hardware component that is found in virtually all computing devices. However, due to its design, which favors serial execution, it is not particularly efficient for highly parallel tasks such as the DNNs [4].

In contrast, GPUs are designed to specialize in image and graphics-related computations across a wide range of devices, including personal computers, workstations, game consoles, and even mobile devices [5]. They are well-suited for DNNs due to their highly parallel architecture, which is particularly effective for the matrix and vector operations frequently encountered in neural network computations. Furthermore, the available robust software ecosystem has greatly accelerated the adoption of these devices [6]. GPUs are mostly used in data centers, where their high computational speed outweighs concerns about the power consumption, as they consume vast amounts of energy [7]. The high energy consumption becomes an issue in cases when DNN applications need to be deployed at the edge, which requires both a small physical size of the device and efficient power management. To address these challenges, researchers have turned towards more specialized hardware that overcomes these issues, even if it does not match GPUs in raw computational power.

Among the specialized hardware solutions, FPGAs are frequently mentioned as an example. They are reconfigurable computer chips that can be programmed to implement any digital hardware circuit. This flexibility is made possible because FPGAs consist of programmable blocks that can be interconnected using routing tracks with programmable switches [8]. The hardware's ability to adapt and implement any hardware circuit is certainly valuable today, as DNNs are evolving rapidly and frequently demand new requirements. Compared to GPUs, FPGAs offer the additional advantage of not requiring a connection to an external CPU to receive instructions. A processor can be implemented directly on the FPGA using programmable logic. However, this approach

is typically slow and consumes a significant portion of the available FPGA area. Instead, it is common for modern FPGAs to integrate a system-on-chip (SoC), which provides a built-in processor directly on the FPGA fabric. As stated by Sciangula et al. [9], this integration is enabled by a high-bandwidth communication bus, allowing fast communication between the processor and the FPGA. The main drawback of SoC-based systems is that they are often more expensive than their non-SoC counterparts.

FPGAs are not the only reconfigurable computing devices available on the market. CGRAs represent an efficient hardware architecture consisting of a two-dimensional array of reconfigurable interconnects and processing elements [10]. Compared to FPGAs, CGRAs utilize a coarse-grained design, which results in both advantages and disadvantages. Luo [11] has presented a valuable comparative study that examines the strengths and weaknesses of both platforms. FPGAs achieve higher throughput due to their ability to perform large-scale bit-level parallel processing, enabled by their fine-grained control. In contrast, CGRAs excel in energy and area efficiency, thanks to simpler routing, coarse-grained control, and reduced redundancy. In practice, both hardware platforms can be effectively utilized depending on the application. For regular and structured computations, CGRAs may be preferable, whereas for experimentation or handling irregular DNNs, FPGAs are generally more suitable.

Reconfigurable hardware is highly useful in many scenarios, as it offers rapid development iterations and reduced time to market when supporting different or slightly modified algorithms. On the other hand, ASICs are optimized for specific algorithms or particular computation operations. As a result, they can achieve higher efficiency than FPGA-based accelerators, but at a cost of a much longer development cycle and higher overall costs. More specifically, the development time and complexity of ASIC platforms far exceed embedded GPUs or FPGAs. This is because the co-design of hardware and software requires developers with professional expertise in both domains. Nonetheless, this results in the most efficient platform, making ASICs a suitable choice for commercial and industrial applications [12].

There are several well-known ASICs, among which the Tensor Processing Unit (TPU) is one of the most outstanding examples. It was designed and developed by Google for use in its data centers to specifically accelerate DNN inference. The motivation behind its development was the rapid increase in DNN complexity, which caused data center computational demands to double within a short period of time. Producing a custom ASIC proved to be more cost-effective than relying on CPUs or GPUs for computation. The TPU was designed to be a co-processor on the PCIe I/O bus, allowing it to be plugged into existing servers in a manner similar to GPUs. In terms of performance, the first-generation TPU achieved inference speeds approximately 15x-30x higher than the K80 GPU and the Haswell CPU, which were state-of-the-art hardware at the time [13]. Although the first-generation TPU was introduced in 2015, Google continued to advance the TPU architecture, and the platform has since evolved to the seventh generation, which operates more like a cluster of chips, delivering massive computational power. This progression demonstrates that ASICs can represent the most efficient solution available. However, due to their high cost and long development cycles, they remain relevant to large technology companies rather than academic researchers.

Using a single platform to perform all computations is not the only available approach. Many researchers instead target hardware heterogeneity for DNN acceleration. As explained by Zahran [14], heterogeneous computing is a scheme in which the different computing nodes possess distinct capabilities or ways of executing tasks. In principle, any type of computing node can be combined to form a scheme; however, in practice, most researchers combine GPUs with either ASICs or FPGAs. This choice is motivated by the fact that GPUs can deliver very high computational performance at the cost of high energy requirements, whereas ASICs and FPGAs prioritize efficiency. Therefore, in such a heterogeneous design, the most resource-intensive or most suitable computations can be offloaded to the more efficient platform, thus reducing overall energy consumption and, in many cases, latency. This approach is supported by the work of Carballo-Hernández et al. [15], who demonstrate that a heterogeneous FPGA-GPU platform outperforms pure GPU-based acceleration for classification inference tasks across multiple networks by significantly reducing energy consumption and overall latency.

B. Software optimization techniques

This subsection focuses on techniques aimed at reducing the size and increasing the speed of accelerated DNNs without any significant accuracy loss. Pruning is a technique that removes negligible connections or weights by setting them to zero. This is possible because operands with very small values are unlikely to substantially affect the overall computation outcome. In practice, pruning is performed by defining a threshold value and setting any weight below this threshold to a numerical zero. Pruning is only beneficial if the target hardware is capable of skipping zero-valued multiplications, since even multiplication by zero requires a clock cycle. Furthermore, higher speeds can be achieved by exploiting the resulting sparsity through encoding zero values so they occupy less storage space. Several encoding techniques exist, ranging from simple run-length coding to compressed sparse column (CSC) and compressed sparse row (CSR). A drawback of these encoding techniques is the requirement for additional encoding and decoding hardware modules [7].

Instead of relying solely on threshold-based pruning, more advanced techniques have been proposed. Petri et al. [16] introduce a novel approach that systematically analyzes network weights and selects a combination that maximizes power efficiency while maintaining sufficient computational accuracy. Their experimental results demonstrate power consumption savings of up to 73.9% in certain cases.

Quantization is another important software-level technique for optimizing weight storage. The most common approach involves reducing weight precision by decreasing the bit-width, for example, from 64 bits to 32 or 16 bits [4]. Traditionally, DNN accelerators have relied on fixed-precision arithmetic, typically using 16-bit or 32-bit floating-point representations. However, it has been shown that lower bit-width formats can significantly reduce computational complexity and memory footprint with minimal impact on model accuracy [17]. Nevertheless, overly aggressive quantization can lead to significant accuracy degradation; therefore, it is important to determine an acceptable accuracy loss and balance it accordingly.

The most widely used quantization method is uniform quantization, which defines equidistant quantization levels

for a given set of weights. In contrast, power-of-two quantization guarantees a higher density of quantization levels around zero, better matching the typical weight distribution in DNNs, with increasing gaps between levels. This approach is primarily well-suited for specialized hardware such as FPGAs and ASICs, as it enables the replacement of traditional multiply-and-accumulate (MAC) units with bitshift-and-accumulate (BAC) units. On general-purpose hardware, uniform quantization may be more effective, as it follows standard arithmetic implementations [18].

Knowledge distillation is a model compression technique that produces a smaller model by transferring knowledge from a larger DL model. This approach uses two models, where the larger network serves as the teacher model and the smaller network acts as the student model. Knowledge transfer is achieved by training the student model to mimic the behavior of the teacher through the minimization of a loss function. This technique is notably effective, as it enables the deployment of high-performing models on resource-constrained embedded devices where the teacher model would not fit otherwise. As a good example of knowledge distillation, the BERT model has been used as a teacher to create the TinyBERT student model. TinyBERT maintains at least 96.8% of the teacher's performance while achieving a 7.5x reduction in model size and 9.4x inference speedup [19].

C. Hardware execution paradigms and architectures

Different hardware platforms adopt distinct execution models and architectural paradigms to exploit parallelism in DNN workloads. General-purpose architectures are typically categorized as temporal architectures, in which parallelism is exploited through techniques such as Single Instruction Multiple Threads (SIMT) or Single Instruction Multiple Data (SIMD). In these temporal designs, arithmetic logic units (ALUs) access data exclusively through the memory hierarchy and cannot communicate directly with one another [20].

In contrast, more specialized hardware platforms, such as ASICs and FPGAs, adopt a spatial architecture that allows each ALU to have its own local memory and control logic, thereby forming a processing element (PE). Depending on their organization, PEs are capable of directly communicating and exchanging data with neighboring elements [21].

These PEs can be interconnected using several communication patterns, including unicast, systolic, multicast, and reduction tree structures. In a unicast configuration, each PE is directly connected to the on-chip buffer, preventing data reuse across PEs. In a systolic arrangement, only the first PE connects to the on-chip buffer, while subsequent PEs are connected sequentially, enabling data to propagate through the array. In a multicast configuration, all PEs independently receive data from the on-chip buffer. Finally, in a reduction tree structure, partial results produced by multiple PEs are combined through a tree-based reduction to form the final output [22].

In spatial architectures storing data closer to the ALU is really efficient as the access is fast and doesn't consume too much energy. Even though storing data like this is favored, these memories are small and thus there is a need to benefit from data reuse as much as possible. Capra et al. [23] describe several commonly used dataflow strategies:

- Weight Stationary: Weights are kept fixed and stored in the register files of the PEs, whereas the inputs and partial sums are distributed across them. This strategy exploits weight and convolutional reuse.
- Output Stationary: Partial sums remain local within a PE, where accumulation proceeds until the final output is produced. During this process, weights and inputs are streamed through the PEs.
- Row Stationary: Convolution operations corresponding to a row of the convolution window are mapped to the same PE, with weights stored in the local register file. This dataflow balances the reuse of input activations, weights, and partial sums.
- No Local Reuse: This dataflow reduces accelerator area by eliminating register files within the PEs, resulting in no stationary data at the PE level.

III. Problem Identification and Future Directions

The progress in hardware acceleration for DNNs has been substantial; several avenues remain open for further research. Although new architectures or even entirely new hardware platforms may emerge in the future, meaningful progress at present depends on targeting existing hardware platforms. Focusing on improving the current architectures is essential for advancing the state-of-the-art.

GPUs have become the dominant platform for DNN training and, in many cases, inference. However, as discussed earlier, their energy consumption remains relatively high. As model complexity continues to increase and societal attention to environmental sustainability grows, further improvements in GPU energy efficiency will be important for their long-term viability. While academic researchers can contribute through algorithmic and software-level optimizations, architectural and low-level innovations are driven by industry, as GPU designs and toolchains are proprietary. Similarly, ASIC-based solutions are associated with a high cost of entry, as their development cycles are both lengthy and expensive. Therefore, future research efforts are better directed toward reconfigurable hardware platforms such as FPGAs and CGRAs, whose flexibility and support for rapid prototyping align well with the goals and constraints of academic research.

When considering CGRA-based accelerators, several promising directions for future research appear. One important area is the development of improved software tools for CGRAs. Such tools could enable more efficient programming of these architectures, thereby lowering the entry barrier for usage. This could encourage wider use of CGRAs, as researchers currently tend to favor FPGAs without fully considering CGRA-based alternatives. Another potential research direction is the exploration of in-memory processing for CGRAs. Even though current CGRA designs employ near-memory processing, data must still be transferred from memory to the computational units. By integrating in-memory processing capabilities, the data movement bottleneck could be reduced, potentially leading to improvements in performance and energy efficiency [20].

Research on FPGA-based acceleration remains an active and important area, with promising directions for future work. One such direction is the development of dynamically reconfigurable hybrid architectures, focusing on FPGA platforms capable of supporting multiple DNN models on a single device. At present, FPGAs are typically programmed for a specific model or a limited set of models. Addressing this

limitation could involve designing hybrid mapping mechanisms that dynamically switch dataflow strategies based on layer characteristics, thereby improving compatibility across a wider range of DNN architectures. Another key research area involves mitigating memory and bandwidth bottlenecks in FPGA-based systems. This could be achieved by designing an optimized memory hierarchy that exploits multi-level memory resources such as high-bandwidth memory (HBM), ultra RAM (URAM), and block RAM (BRAM), or by introducing a hierarchical cache architecture that would significantly reduce off-chip memory access and enhance on-chip storage utilization. In addition, refining high-level synthesis (HLS) technology to support direct generation of high-performance hardware from high-level languages such as C++ remains an open challenge. Although such generation is currently possible, achieving efficient designs still requires significant manual tuning and substantial hardware expertise [24].

Future research does not need to be limited to a single hardware platform. In fact, a growing trend is the exploration of heterogeneous systems that integrate multiple types of processors. This approach is well motivated, as different platforms offer complementary advantages, allowing each processor to focus on the tasks for which it is most efficient while offloading less suitable workloads to other components [25].

IV. Conclusion

Hardware acceleration of DNNs has become increasingly important due to the rapidly growing complexity of modern models. Acceleration can be achieved using both general-purpose platforms, such as CPUs and GPUs, and specialized hardware, including FPGAs, CGRAs, and ASICs. Selecting the appropriate hardware platform enables more efficient execution of target DNN workloads. Furthermore, both hardware- and software-level optimization techniques can be utilized to improve performance and efficiency.

Although GPUs and ASICs are widely used for DNN acceleration, their high development costs and long design cycles present significant barriers to entry. As a result, reconfigurable platforms such as FPGAs and CGRAs, along with heterogeneous systems that combine multiple hardware architectures, represent promising directions for future research.

Acknowledgment

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

References

[1] O. S. Ekundayo and A. E. Ezugwu, "Deep learning: Historical overview from inception to actualization, models, applications and future trends," *Applied Soft Computing*, vol. 181, p. 113378, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494625006891

[2] V. Sze, Y. Chen, T. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *CoRR*, vol. abs/1703.09039, 2017. [Online]. Available: http://arxiv.org/abs/1703.09039

[3] Y. Shi, "Unveiling the powerhouses of AI: A comprehensive study of GPU, FPGA, and ASIC accelerators," vol. 50, pp. 97–105. [Online]. Available: https://ace.ewapub.com/article/view/10910

[4] H. Wang, X. Liu, Y. Yang, Q. Liao, S. Ke, and Y. Dong, "FPGA-based acceleration of deep learning networks: Techniques and applications," in *Proceedings of the 2025 6th International Conference on Computer Information and Big Data Applications*, ser. CIBDA '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 774–778. [Online]. Available: https://doi.org/10.1145/3746709.3746843

[5] Y. Hu, Y. Liu, and Z. Liu, "A survey on convolutional neural network accelerators: GPU, FPGA and ASIC," in *2022 14th International Conference on Computer Research and Development (ICCRD)*, 2022, pp. 100–107.

[6] R. Bishwokarma and X. Xiong, "Accelerating deep neural networks in FPGA SoC," in *2024 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, 2024, pp. 1–7.

[7] R. Machupalli, M. Hossain, and M. Mandal, "Review of ASIC accelerators for deep neural network," *Microprocessors and Microsystems*, vol. 89, p. 104441, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141933122000163

[8] A. Boutros and V. Betz, "FPGA architecture: Principles and progression," *IEEE Circuits and Systems Magazine*, vol. 21, no. 2, pp. 4–29, 2021.

[9] G. Sciangula, F. Restuccia, A. Biondi, and G. Buttazzo, "Hardware acceleration of deep neural networks for autonomous driving on FPGA-based SoC," in *2022 25th Euromicro Conference on Digital System Design (DSD)*, 2022, pp. 406–414.

[10] N. Li, D. Li, Z. Wu, P. Jing, S. Hang Pun, and Y. Liu, "Submap: A partial mapping strategy for CGRA based on sub-CGRA exploration," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 44, no. 7, pp. 2827–2831, 2025.

[11] J. Luo, "A comparative study of FPGA and CGRA technologies in hardware acceleration for deep learning," vol. 51, pp. 168–180. [Online]. Available: https://ace.ewapub.com/article/view/10883

[12] R. Wu, X. Guo, J. Du, and J. Li, "Accelerating neural network inference on FPGA-based platforms—a survey," *Electronics*, vol. 10, no. 9, 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/9/1025

[13] N. P. Jouppi *et al.*, "In-datacenter performance analysis of a tensor processing unit," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1–12. [Online]. Available: https://doi.org/10.1145/3079856.3080246

[14] M. Zahran, "Heterogeneous computing: here to stay," *Commun. ACM*, vol. 60, no. 3, p. 42–45, Feb. 2017. [Online]. Available: https://doi.org/10.1145/3024918

[15] W. Carballo-Hernández, M. Pelcat, and F. Berry, "Why is FPGA-GPU heterogeneity the best option for embedded deep neural networks?" 2021. [Online]. Available: https://arxiv.org/abs/2102.01343

[16] R. Petri, G. L. Zhang, Y. Chen, U. Schlichtmann, and B. Li, "Powerpruning: Selecting weights and activations for power-efficient neural network acceleration," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.

[17] Muhammed Raees P. C., Akshayraj M. R., V. P. Gopi, G. Lakshminarayanan, G. R. Gangadharan, and J. U. Kidav, "Dynamic precision scaling in MAC units for energy-efficient computations in deep neural network accelerators," in *2024 28th International Symposium on VLSI Design and Test (VDAT)*, 2024, pp. 1–7.

[18] D. Przewlocka-Rus and T. Kryjak, *Energy Efficient Hardware Acceleration of Neural Networks with Power-of-Two Quantisation*. Springer Nature Switzerland, 2023, p. 225–236. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-22025-8_16

[19] M. M. H. Shuvo, S. K. Islam, J. Cheng, and B. I. Morshed, "Efficient acceleration of deep learning inference on resource-constrained edge devices: A review," *Proceedings of the IEEE*, vol. 111, no. 1, pp. 42–91, 2023.

[20] P. Dhilleswararao, S. Boppu, M. S. Manikandan, and L. R. Cenkeramaddi, "Efficient hardware architectures for accelerating deep neural networks: Survey," *IEEE Access*, vol. 10, pp. 131 788–131 828, 2022.

[21] D. Ghimire, D. Kil, and S.-h. Kim, "A survey on efficient convolutional neural networks and hardware acceleration," *Electronics*, vol. 11, no. 6, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/6/945

[22] L. Jia, Z. Luo, L. Lu, and Y. Liang, "Automatic generation of spatial accelerator for tensor algebra," *Ieee Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, pp. 1898–1911, 2023.

[23] M. Capra, B. Bussolino, A. Marchisio, M. Shafique, G. Masera, and M. Martina, "An updated survey of efficient hardware architectures for accelerating deep convolutional neural networks," vol. 12, no. 7, p. 113. [Online]. Available: https://www.mdpi.com/1999-5903/12/7/113

[24] L. Gao, Z. Luo, and L. Wang, "Convolutional neural network acceleration techniques based on FPGA platforms: Principles, methods, and challenges," *Information*, vol. 16, no. 10, 2025. [Online]. Available: https://www.mdpi.com/2078-2489/16/10/914

[25] B. J. Kang, H. I. Lee, S. K. Yoon, Y. C. Kim, S. B. Jeong, S. J. O, and H. Kim, "A survey of FPGA and ASIC designs for transformer inference acceleration and optimization," *Journal of Systems Architecture*, vol. 155, p. 103247, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S138376212400184X

Augmentation method for robust UAV-based military object detection

[1]*Ladislav Hric (2nd year),*
Supervisor: [2]*Daniela Perduková*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Košice, Slovak Republic

[1]ladislav.hric@tuke.sk, [2]daniela.perdukova@tuke.sk

***Abstract*—The growing use of unmanned aerial vehicles (UAVs) in military operations challenges object detection due to occlusions, camouflage, and varying viewing angles, while computational limits constrain onboard deployment. This work addresses these challenges by investigating an occlusion-focused data augmentation technique for military object detection from UAV imagery. A custom augmentation algorithm, leveraging the Segment Anything Model (SAM), simulates partial occlusion by inserting local samples from object surroundings. Experiments were conducted on a manually annotated dataset of 4,987 images across five object classes, along with a specialized test set of 175 heavily masked instances to evaluate model robustness. The baseline YOLOv8 model was compared with a model trained on the augmented dataset, yielding a 13.9% improvement in overall $\mathbf{mAP}_{50-95}$. These results highlight the augmentation's potential and encourage further evaluation.**

***Keywords*—augmentation, dataset, object detection, unmanned aerial vehicle.**

I. Introduction

The current geopolitical tension and the ongoing armed conflict in Ukraine, often referred to as the "drone war," highlight the growing importance of unmanned aerial vehicles (UAVs) in modern military operations. UAVs are used not only for offensive tasks, but also for defensive operations, reconnaissance, and surveillance, providing rapid and flexible situational awareness in areas of interest [1], [2], [3].

One of the key challenges in designing UAV detection systems is achieving sufficient accuracy under limited computational resources and challenging visual conditions, such as object occlusion, poor lighting, or variations in object viewpoints caused by aerial perspectives that differ from traditional ground-based detection. In military applications, the situation is even more challenging, as target objects are often intentionally camouflaged. Another major challenge is the difficulty of obtaining samples of target objects across all possible scenarios and conditions that may occur during real UAV operations. These factors significantly affect the ability of detection models to generalize to real-world scenarios. Therefore, data augmentation is particularly useful, as it allows existing datasets to be expanded with more challenging samples and simulates diverse visual situations without the need for complex new data collection.

The motivation of this work is to contribute to addressing these challenges through research on the detection of military objects from UAV imagery, focusing on improving the robustness of detection models under conditions of visual camouflage. This contribution summarizes the progress achieved in the dissertation research and provides a foundation for its further development.

II. Background

Most studies in the field of object detection focus on ground-based imagery acquired from conventional cameras or stationary devices. However, UAV-based detection presents specific challenges, such as changes in object scale depending on flight altitude, varying viewing angles, UAV movement, and complex environments, which significantly affect the accuracy of detectors [4]. Deploying detection models directly onboard a UAV also requires consideration of limited computational resources. Model optimization and the use of smaller networks enable real-time processing while maintaining relatively good accuracy.

Several architectures can be used, including two-shot detectors such as Fast R-CNN [5], or single-shot detectors such as SSD [6] or YOLO [7]. Recently, Vision Transformers [8] have gained popularity; however, they are generally considered computationally more demanding, which may limit their deployment on UAVs.

Research focused on military applications faces significant limitations. One of the biggest challenges is the lack of high-quality datasets containing military equipment, particularly from the perspective of UAV and in various scenarios [9], [10]. Although there are some well-known public datasets, such as COCO [11], that are commonly used to compare model performance, these are mainly focused on civilian applications and only a few contain UAV imagery. Therefore, many studies create their own datasets from online sources or using simulations or physical models.

The lack of training data significantly affects the quality of the detection output. Data augmentation techniques are useful for addressing this issue, as they expand the original dataset with new training samples. Various approaches exist, including transformation-based techniques that change object orientation, scale, or perspective (e.g., rotation, scaling, horizontal flipping), or color-based techniques that adjust brightness, contrast or saturation [12]. When it is necessary for the model to robustly detect partially occluded objects, occlusion-focused augmentation methods are applied – a typical example is Cutout, where parts of the image are randomly masked [13].

Together, these approaches enable the development of detection systems capable of reliably identifying objects from UAV imagery, even with limited data and under challenging real-world conditions.

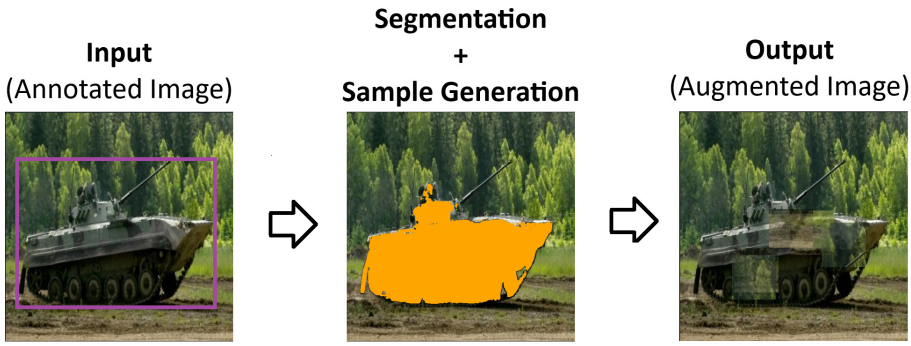


Fig. 1: Process of the Proposed Augmentation .

TABLE I: Experimental results of the reference model

Class	Images	Instances	Precision	Recall	mAP_{50-95}
Total	108	175	0.516	0.229	0.257
Airplane	27	62	0.688	0.177	0.256
Helicopter	27	32	1.000	0.438	0.499
Tank	38	64	0.895	0.531	0.532
UAV	8	8	0.000	0.000	0.000
Vehicle	9	9	0.000	0.000	0.000

TABLE II: Experimental results of the model with the proposed augmentation technique

Class	Images	Instances	Precision	Recall	mAP_{50-95}
Total	108	175	0.696	0.290	0.396
Airplane	27	62	0.632	0.194	0.314
Helicopter	27	32	1.000	0.406	0.530
Tank	38	64	0.846	0.516	0.545
UAV	8	8	0.000	0.000	0.000
Vehicle	9	9	1.000	0.333	0.589

III. Recent Achievements

In the recent period, the experimental part of the thesis was conducted to evaluate the potential of the proposed augmentation technique to improve the precision and generalization capabilities of detection models when identifying partially occluded or masked military objects in images captured by UAVs. The experiments represent a baseline phase for further research and development and do not constitute a final solution.

Datasets

First, an initial version of the training dataset consisting of 4,987 images was created of military objects. The data were manually annotated and divided into training, validation, and test sets with a ratio of 86 % / 11 % / 3 %.

The dataset contains five classes of military equipment with the following numbers of annotated instances:

- Aircraft – 2,273
- Helicopter – 408
- Tank – 3,456
- UAV – 2,131
- Vehicle – 1,204

This dataset represents a baseline reference set containing predominantly standard visual scenarios with fully or partially visible objects. It serves as a starting point for training detection models and enables objective comparison with models trained on augmented data. Future work aims to improve class balance; however, for the purposes of initial experiments, this limitation can be considered negligible.

In addition to the baseline dataset, a separate test set was constructed, specifically focused on 175 heavily masked military objects. It contains no samples from the original data collection and was designed to emphasize challenging real-world scenarios that are rarely represented in common public datasets. The selected samples depict equipment concealed either directly through camouflage colors and textures or indirectly by foreign objects such as tree branches, dense vegetation, netting, or other scene obstructions. During the construction of this dataset, a limitation was encountered in the availability of heavily masked instances of the UAV class. However, this reflects real-world conditions, as UAVs are less prevalent than ground vehicles and typically require less visual camouflage during short-term aerial deployment, particularly due to their design and size characteristics.

Reference model

The reference detection model was based on the YOLOv8 architecture, which provides a favorable compromise between detection accuracy and computational efficiency, a key requirement for onboard deployment on UAV platforms. The model was trained for 100 epochs with a batch size of 16. Quantitative evaluation results obtained on the masked-object test dataset are summarized in Table I. The model achieved an overall mAP_{50-95} of 0.257. Performance for the UAV and Vehicle classes was zero, indicating the limited capability of the model to detect heavily masked objects. The best results were obtained for the Tank class, which can be attributed to the larger number of training instances and the higher number of relevant test samples.

Augmentation

To improve performance, a custom augmentation algorithm was designed and implemented, simulating partial occlusion of objects using visual elements from their surroundings. The process (Fig. 1) utilizes the Segment Anything Model (SAM) [14] for object segmentation and randomly inserts local samples from the surrounding area directly onto the object. Augmentation parameters include the number of patches, patch size, and transparency.

Initially, the random parameters were set as follows: number of patches = 5, patch size = 30% of the lower annotation dimension, and transparency = 0.6. The model trained on the augmented dataset showed a significant improvement: overall mAP_{50-95} increased to 0.396, representing a 13.9% improvement over the reference model. As summarized in Table II , the largest gains were observed for the Aircraft, Helicopter, and Vehicle classes, while performance for the UAV class remained low. Inference time was measured at approximately 7 ms per image; however, it should be noted that the tests were performed on a Dell Inspiron 15 7000 notebook.

Overall, the experiments and obtained results represent a significant step in the development of augmentation methods for object detection and provide a foundation for future model optimization, dataset expansion, and real-world testing.

IV. Future Plans

The results demonstrate that the proposed augmentation technique has the potential to improve model performance in detecting heavily masked objects, which is particularly relevant for UAV-based object detection in military applications. To confirm these findings and provide a more robust evaluation, the test dataset will be extended, constituting the first step in the planned future work. The second step will involve determining the optimal parameters for the dataset. To further assess the effectiveness of the augmentation method, it should be compared with relevant alternative approaches, representing the third step of future research. Based on the results, it may be considered to modify the current approach to achieve better performance or to simplify the technique. The final planned step is to integrate the system into the UAV platform and evaluate it under real operational constraints and conditions that simulate military scenarios.

Acknowledgment

This work was supported by Ministry of Defence of the Slovak Republic under the Contract no. SEMOD-EL76/93-11/2025.

References

[1] B. Zeng, S. Gao, Y. Xu, Z. Zhang, F. Li, and C. Wang, "Detection of military targets on ground and sea by uavs with low-altitude oblique perspective," *Remote Sensing*, vol. 16, no. 7, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/7/1288

[2] L. Yue, H. Ling, J. Yuan, and L. Bai, "A lightweight border patrol object detection network for edge devices," *Electronics*, vol. 11, no. 22, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/22/3828

[3] Z. Xiaoning, "Analysis of military application of uav swarm technology," in *2020 3rd International Conference on Unmanned Systems (ICUS)*, 2020, pp. 1200–1204.

[4] G. Tang, J. Ni, Y. Zhao, Y. Gu, and W. Cao, "A survey of object detection for uavs based on deep learning," *Remote Sensing*, vol. 16, no. 1, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/1/149

[5] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *CoRR*, vol. abs/2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[9] F. Kamran, M. Shahzad, and F. Shafait, "Automated military vehicle detection from low-altitude aerial images," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018, pp. 1–8.

[10] H. Liu, Y. Yu, S. Liu, and W. Wang, "A military object detection model of uav reconnaissance image and feature visualization," *Applied Sciences*, vol. 12, no. 23, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/23/12236

[11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[12] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C. C. Aggarwal, J. Pei, and Y. Zhou, "A comprehensive survey on data augmentation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 38, no. 1, pp. 47–66, 2026.

[13] T. Devries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *CoRR*, vol. abs/1708.04552, 2017. [Online]. Available: http://arxiv.org/abs/1708.04552

[14] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.

Weather-Based Spatial Dynamic Line Rating of Overhead Transmission Lines Using a CIGRE 601 Thermal Model

[1]*František Margita (3rd year),*
Supervisor: [2]*Ľubomír Beňa*

[1,2]Department of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]frantisek.margita@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract*—This paper presents a spatial dynamic line rating (DLR) model based on ampacity tracing using meteorological conditions, geographical line points, and a thermal model of an overhead transmission line (OHTL), as defined in CIGRE Technical Brochure 601 (CIGRE 601). The proposed approach identifies the limiting point for ampacity, thereby ensuring a safe and realistic determination of ampacity with respect to the entire line. The methodology was applied to a 400 kV OHTL under Central European conditions and was compared with a conservative static line rating (SLR). The results showed that spatial DLR provides a consistent basis for the application of weather-based DLR and represents a suitable technical, economic, and operational framework for extending ampacity assessment in the Slovak transmission system (TS).**

***Keywords*—ampacity, CIGRE 601, forecasted weather conditions, spatial dynamic line rating**

I. Introduction and Motivation

From a practical perspective, a reliable and safe DLR implementation requires the deployment of sensors or advanced methods capable of identifying critical sections based on a thermal and mechanical model of an OHTL. Therefore, the motivation is to design a spatial DLR approach that monitors meteorological conditions, conductor temperature, and the resulting ampacity along the entire OHTL.

Spatial DLR enables the calculation of realistic ampacity for the entire OHTL and represents a promising solution in terms of operational safety, technical feasibility, and economic benefits. With a sufficiently high-quality and representative sample of meteorological data, spatial DLR can also be applied to seasonal DLR analysis, the identification of critical line sections, and decision support for logistics and optimized deployment of sensors for other DLR implementations.

II. Background and Research Progress

The last two academic years of the doctoral study were focused on the following research activities:

- the development of a research framework and overview of available approaches based on CIGRE 601, including a comparison with IEEE 738 for the calculation of the thermal model and ampacity of OHTLs,
- the calculation of SLR under the conditions of the Slovak TS in accordance with the STN EN 50341-2-23 standard,
- the calculation of steady-state DLR and its comparison with conservative SLR,
- the assessment of high-temperature low-sag aluminum conductor composite core (HTLS, ACCC) conductors as a replacement for conventional ACSR conductors, including the evaluation of their benefits using the proposed thermal models,
- the application of the proposed SLR and steady-state DLR models to the 400 kV V427 OHTL (Rimavská Sobota – Moldava), including the use of alternative HTLS ACCC conductors with the aim of increasing ampacity,
- the development of a numerical model for transient DLR based on the transient thermal equation defined in CIGRE 601,
- the investigation of the effects of emissivity and absorptivity changes due to conductor aging on ampacity,
- the experimental verification of the proposed thermal models under real conditions in the laboratory of the Lublin University of Technology.

III. Recent Achievements

A. Key Research Achievement

A major research achievement was the successful development of a transient DLR model, which was subsequently verified during an Erasmus+ research stay at the Lublin University of Technology. The proposed methodology and the interpretation of the results were published in a peer-reviewed Q1 journal, *International Journal of Electrical Power & Energy Systems* [1].

At the end of 2025, the article entitled *Transient dynamic ampacity of OHTLs utilizing CIGRE 601: Application under conditions in the Slovak Republic* was awarded by the Rector of the Technical University of Košice in the competition *Publication Activity of Doctoral Students 2025* [2].

B. Tasks Completed in the Previous Year

A major milestone in fulfilling the dissertation objectives and in advancing the approach to ampacity assessment was the calculation of steady-state DLR using the proposed line-tracing method. This method consisted of the following components:

- the development of an algorithm for processing geometric coordinates of OHTLs within the Slovak TS,

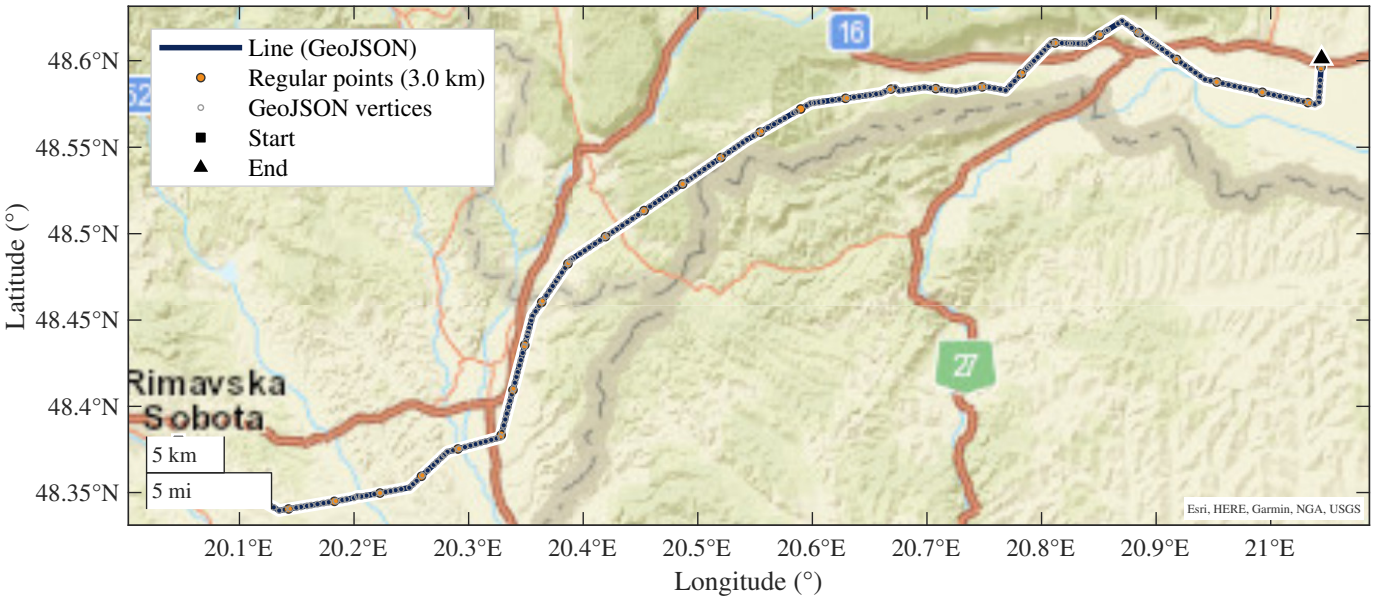


Fig. 1. Map of the 400 kV Rimavská Sobota – Moldava nad Bodvou OHTL generated from the obtained geographical coordinates

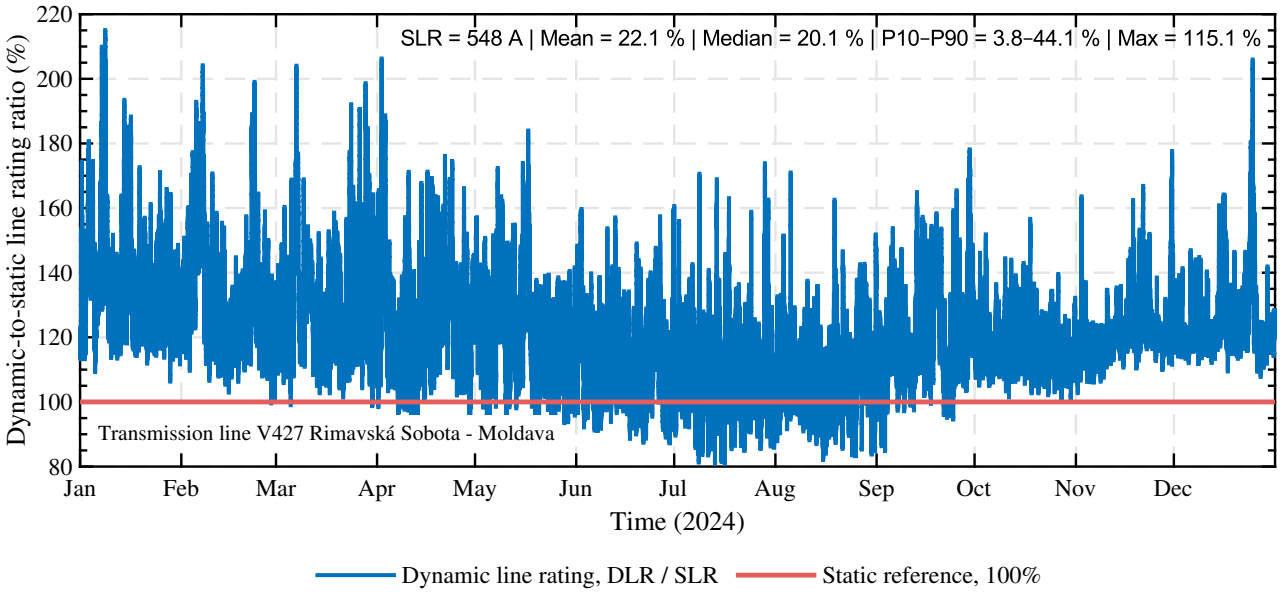


Fig. 2. Year-2024 spatial DLR profile expressed as the DLR/SLR ratio at 15-min resolution

- the development of an algorithm for reconstructing and visualizing OHTL routes from the acquired geometric coordinates (Fig. 1),
- the development of an algorithm for acquiring meteorological data for the reconstructed line coordinates with a 15-minute temporal resolution over an entire year or a specified time interval (Fig. 2),
- the development of an algorithm for calculating steady-state DLR from the acquired data in order to determine safe ampacity along the entire OHTL at the maximum operating temperature of the selected conductor.

From a physical perspective and with respect to dynamic ampacity accuracy, the model was extended compared to the previous version by implementing automatic altitude calculation and by introducing the conversion of conductor resistance from direct current resistance R_{DC} to alternating current resistance R_{AC}.

From an operational and user-oriented perspective, the model was fully automated and parameterized. This enabled a more efficient interpretation of the results for different conductor types, various 400 kV OHTLs, and systematic processing of meteorological database inputs.

During the previous year, the developed methodology and the results obtained using the proposed model were prepared for publication in a Q1 journal in the field of electric power engineering.

The tracing algorithm enabled the construction of a time-dependent ampacity profile along the entire OHTL, thereby providing a physically consistent basis for operational and planning analyses.

IV. Results

Figure 2 shows the annual ampacity profile expressed as the ratio of steady-state DLR (blue line) to SLR (red line). The

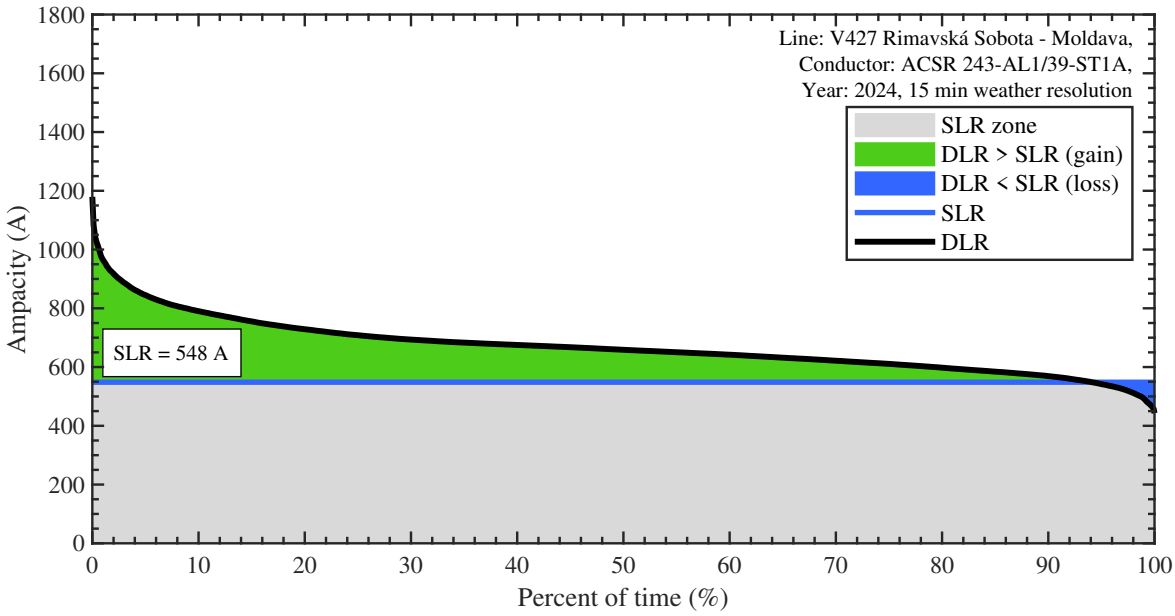


Fig. 3. Percentage of ampacity during 2024 using spatial DLR, based on 15-minute operating parameters

SLR for the analyzed conductor was calculated in accordance with the applicable standard (STN EN 50341-2-23) as 548 A, corresponding to a reference value of 100 %.

Steady-state DLR exceeded SLR in 94.11 % of the evaluated cases, corresponding to 8265.8 hours or 344 days in 2024. The mean DLR level reached 122.1 % (670 A), with a median value of 120.1 % (659 A). Percentile analysis indicated that the 10th and 90th percentiles were 103.8 % and 144.1 %, respectively. The maximum recorded ampacity reached 215.1 % of SLR (1179 A). Conversely, DLR values below SLR occurred in 5.89 % of cases, corresponding to 22 days, with a minimum ampacity of 443 A.

Seasonal analysis revealed that the least favorable ampacity conditions were observed in August, with an average DLR of 107.6 % (590 A), while the highest ampacity levels were recorded in January, reaching 136.6 % (749 A).

Based on Fig. 3, percentile analysis showed that a DLR of 659 A was available for 50 % of the operational time (120 % of SLR), whereas a DLR of 611 A was available for 75 % of the time (112 % of SLR) during line operation in 2024.

V. Discussion

By comparing the outputs presented in Fig. 1 with a map of the Slovak TS [3], it was confirmed that the coordinate acquisition algorithm is correct and can be reliably applied in the calculation of spatial DLR.

The analysis of the DLR-to-SLR ratio indicated that the SLR is dimensioned conservatively under the evaluated operating conditions.

Furthermore, the results highlighted a pronounced seasonal dependence of ampacity on meteorological conditions, which remains unexploited when line operation is constrained by a single SLR applied uniformly throughout the year.

According to [4], many OHTLs can safely apply DLR up to 130 % of their static ampacity for a substantial portion of the year. In contrast, under Central European conditions and identical operating assumptions (130 % of SLR, 713 A), the analyzed OHTL reached this ampacity level during only 23.72 % of the valid annual profile in 2024. This discrepancy can be attributed to the high variability of regional meteorological conditions, particularly changes in wind speed and direction, the duration of the data acquisition period, and differences in the methodologies used for the calculation of convective cooling [5].

VI. Future Plans

Future research will focus on completing the doctoral dissertation, specifically on a safe normative increase in the operating temperature of conductors rated for a lower temperature, the design of software for calculating SLR and DLR, the economic evaluation of the proposed solutions, including the identification of their benefits, and the formulation of recommendations for practice.

Acknowledgment

This paper was supported by the Science Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under the contract VEGA 1/0532/25.

References

[1] F. Margita, Ľ. Beňa, T. Binkowski, P. Pijarski, and S. Adamek, "Transient dynamic ampacity of overhead transmission lines utilizing cigre technical brochure 601: Application under conditions in the slovak republic," *International Journal of Electrical Power & Energy Systems*, vol. 171, p. 110941, 2025.

[2] Technical University of Košice, "Results of the "publication activity of doctoral students" competition 2025," TUKE News, Dec. 2025, published December 4, 2025. [Online]. Available: https://tuke.sk/sk/archive/clanky/vysledky-sutaze-publikacna-cinnost-doktorandov-v-roku-2025

[3] Slovenská elektrizačná prenosová sústava, a.s., "Schemes of the slovak transmission system," Online, 2025, accessed January 2026. [Online]. Available: https://www.sepsas.sk/pre-partnerov/technicke-parametre/schemy-siete/

[4] N. H. Abas, M. Z. A. Ab Kadir, N. Azis, J. Jasni, N. F. Ab Aziz, and Z. M. Khurshid, "Optimizing grid with dynamic line rating of conductors: A comprehensive review," *IEEE Access*, vol. 12, pp. 9738–9756, 2024.

[5] Y. Wang, F. Fan, Y. Wang, K. Wang, J. Jiang, C. Sun, R. Xue, and K. Song, "Convective heat loss prediction using the concept of effective wind speed for dynamic line rating studies," *Energies*, vol. 18, no. 16, p. 4452, 2025.

Artificial Intelligence for Routing in 6G Mobile Multi-Hop Networks: A Progress Report and Comparative Study

[1]*Jozef Badár (2nd year),*
Supervisor: [2]*Ján Papaj*

[1,2]Dept. of Computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]jozef.badar@tuke.sk, [2]jan.papaj@tuke.sk

***Abstract*—Mobile ad hoc networks (MANETs) represent a major routing challenge in the 5G/6G context, especially under increasing node density and stricter quality-of-service (QoS) requirements. Classical reactive protocols such as AODV show significant performance degradation under high load and mobility. This paper summarizes recent progress in a systematic comparison of classical MANET routing and artificial-intelligence-based approaches implemented in NS-3 3.36.1, including AODV, bio-inspired metaheuristics (ACO, PSO, Hybrid PSO/ACO), and selected deep reinforcement learning methods (DQN, Monte Carlo), complemented by experimental A3C and GNN routing implementations. Experiments for 20–200 mobile nodes under identical conditions evaluate packet delivery ratio (PDR), end-to-end delay, throughput, hop count, and computational cost. Results show that Hybrid PSO/ACO achieves the highest delivery reliability (94.8–99.5%) with low delay (1.16–4.53 ms), while DQN provides the highest throughput in dense scenarios (up to 462 kbps). ACO offers a balanced compromise across metrics. The findings provide practical recommendations for selecting routing mechanisms in 5G/6G MANET deployments with edge-computing support.**

***Keywords*—MANET, routing, NS-3, swarm intelligence, reinforcement learning, 5G, 6G**

I. Introduction and Motivation

MANET routing is critical for 5G/6G systems with dynamic topologies and strict QoS constraints, where classical reactive protocols (AODV, DSR, OLSR) degrade beyond 80 nodes [1], [2]. AI and metaheuristics provide a promising alternative. The motivation is the need for reliable low-latency routing in URLLC, V2X, and edge scenarios. The objective is to systematically compare 8 routing approaches in a unified NS-3 environment.

Hypotheses: (H1) AI/metaheuristic routing achieves higher PDR than AODV above 80 nodes; (H2) Hybrid PSO/ACO minimizes latency; (H3) DQN maximizes throughput in dense scenarios.

II. Background and Progress

In the initial phase, the routing problem in dense 5G/6G MANET scenarios was formally defined, evaluation metrics were specified, and a unified NS-3 experimental framework was prepared. The limitations of classical protocols were analyzed and literature gaps were identified, motivating a comparison of AI and metaheuristic methods.

III. Recent Achievements

A. Completed Research Tasks

5G and 6G architectures emphasize AI-native control and edge computing. MANET multi-hop connectivity is relevant for V2X, UAV swarms, and partially degraded infrastructure [1], [2]. Classical protocols (AODV, DSR, OLSR) are limited under high mobility and density, while AI-based methods (DQN, ACO, PSO, GNN) are promising [3].

The literature reveals critical gaps: (i) comparisons often use different simulators and parameter sets; (ii) Hybrid PSO/ACO is underexplored empirically; (iii) RL hyperparameters are often insufficiently reported; (iv) computational cost on edge devices is frequently ignored; (v) mechanistic WHY explanations are rare. During the evaluated period, these issues were addressed through a unified experiment design and reproducible implementation.

Key achievements: (1) a unified NS-3 benchmark for 8 approaches (20–200 nodes, 5 metrics), (2) implementation of Hybrid PSO/ACO, (3) empirical validation of RL and GNN approaches, (4) computational-cost analysis for edge devices. The results build on the author's publications from the evaluated period [4], [5], [6], [7], which document recognized progress and continuity of the research line.

From the perspective of objective verification, the hypotheses were consistently confirmed across evaluated scenarios. H1 is supported by stable PDR improvements of AI/metaheuristic methods over AODV at higher densities. H2 is supported by the dominant latency performance of Hybrid PSO/ACO versus competing methods. H3 is supported by the highest throughput of DQN in dense topologies. These findings provide a clear methodological basis for transfer to testbed validation.

B. Experimental Platform and Implemented Methods

NS-3 3.36.1 [8] with IEEE 802.11, 20–200 nodes, random waypoint mobility, and CBR/UDP traffic. Implemented methods: AODV, DQN, MC, ACO, PSO, Hybrid PSO/ACO, plus experimental A3C and GNN.

C. Methodology and Implementation

1) Routing Algorithms: Implemented: AODV (reactive reference protocol with RREQ/RREP signaling), ACO (pheromone-based optimization), PSO (swarm optimization), Hybrid

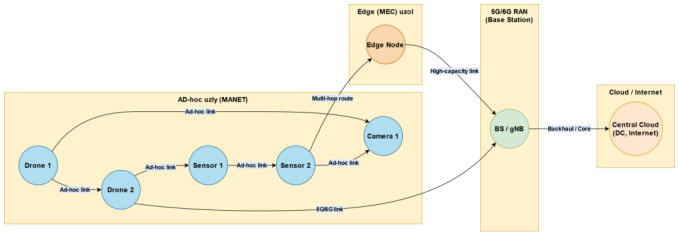


Fig. 1. Schematic representation of a multi-hop topology in a 5G/6G network with edge-node integration.

PSO/ACO (global PSO exploration with local ACO refinement), DQN (deep Q-network, replay buffer 20K, batch 32), Monte Carlo (return-based action estimation), plus experimental A3C (actor-critic, on-policy) and GNN (topology-oriented model with limited expressivity without pretraining). State vector: position and distances ($2N+2$ and $3N+2$). Reward: delivery success with $\gamma = 0{,}95$.

TABLE I
CONFIGURATION OF IMPLEMENTED RL AND GNN ALGORITHMS (N = NUMBER OF NODES)

Parameter	**DQN** [9]	**A3C** [10]	**MC**	**GNN** [11]
Input	$2N+2$	$2N+2$	$2N+2$	$3N+2$
Architecture	$128\to128\to N$	Ac: $128\to N$; Cr: $128\to1$	$128\to128\to N64\to64\to N$	
Activation/output	ReLU/linear	ReLU/soft-max+lin.	ReLU/linear	ReLU/linear
Optimizer (lr)	Adam ($5{\cdot}10^{-4}$)	Adam (10^{-3})	Adam ($5{\cdot}10^{-4}$)	Adam (10^{-3})
γ	0,95	0,95	0,95	–
Replay buffer	20 000	–	20 000	–
Batch	32	on-policy	MC return	–
Training	online	online	per episode	–
Episodes	50	50	50	50

2) Scenarios, Metrics, and Parameters: Experiments: 20–200 nodes, random waypoint mobility, CBR/UDP flow, 50 episodes. Metrics: PDR, delay, throughput, hop count, runtime. Table I reports hyperparameters; all scenarios used fixed RngSeed (12345) on Linux with AMD Ryzen 9 7900X and RTX 4060Ti.

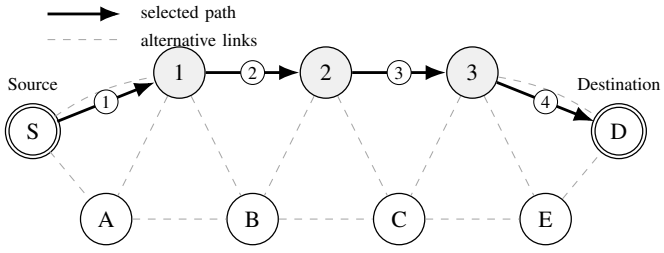


Fig. 2. Example of a multi-hop path between source S and destination D in a MANET. The bold numbered route denotes the selected path, while dashed lighter links represent alternative connections considered by routing agents. Efficient algorithms must select paths with low delay and high reliability under competing options.

D. Quantitative Results and Interpretation

Hybrid PSO/ACO: PDR 94.8–99.5%, delay 1.16–4.53 ms (latency-critical use cases). DQN: throughput 462 kbps at 200 nodes (throughput-critical). ACO: balanced trade-off. AODV: clear degradation above 80 nodes (statistically significant, p<0.05). A3C/GNN without pretraining: weak results. WHY mechanisms: PSO/ACO—parallel search; DQN—pattern learning; ACO—pheromone stability; AODV—high signaling overhead.

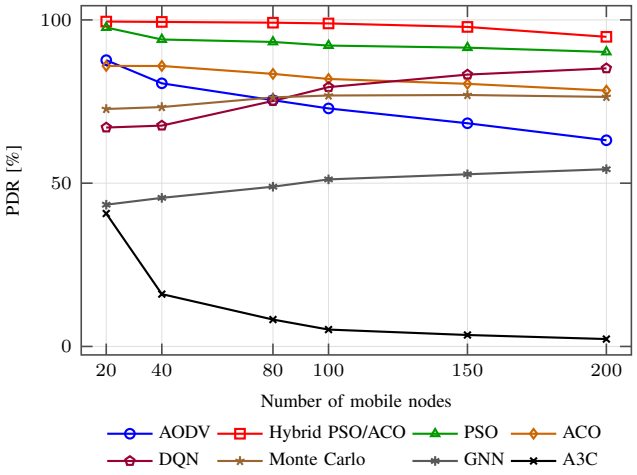


Fig. 3. PDR comparison versus node count. Hybrid PSO/ACO and PSO maintain high reliability up to 200 nodes (95%), while AODV drops from 88% to 63% due to increased signaling overhead and collisions. A3C and GNN show poor performance (40% and below), indicating insufficient convergence in larger scenarios without supervised pretraining.

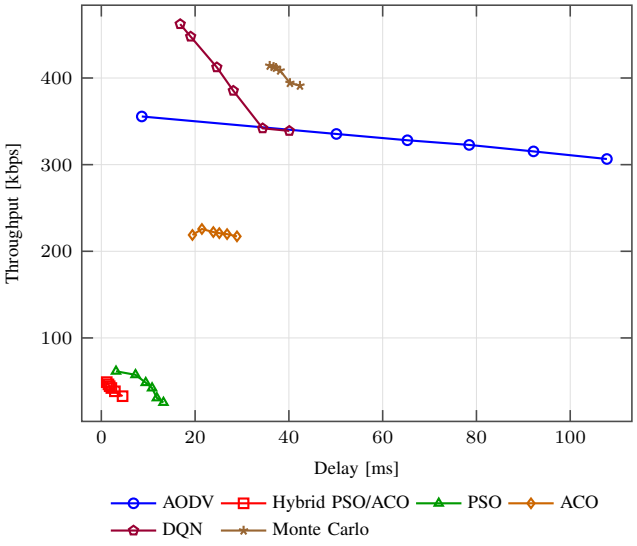


Fig. 4. Delay-throughput trade-off across 6 density scenarios. DQN achieves maximum throughput (462 kbps at 200 nodes) at the cost of higher delay; Hybrid PSO/ACO prioritizes the lowest delays (1–4 ms) with lower throughput (25–50 kbps); ACO and Monte Carlo occupy the middle region with competitive values on both metrics. This Pareto frontier is central for algorithm selection by application requirements.

The results show that swarm-intelligence and RL-based routing can significantly outperform classical reactive MANET protocols in a consistent NS-3 framework [2], [3]. Mechanistic explanations for method performance are as follows:

Hybrid PSO/ACO dominance in PDR and delay: Combining global PSO exploration with local ACO pheromone refinement creates synergy. PSO quickly converges to high-quality path candidates, while ACO strengthens and stabilizes reliable routes. As a result, PDR remains at 95–99% with 1–5 ms delay, compared to AODV degradation from 88% to 63%.

DQN throughput: Q-learning adaptively avoids congested nodes, at the cost of higher delay. *ACO stability:* Pheromone dynamics converge robustly in lower-density regimes. *A3C,*

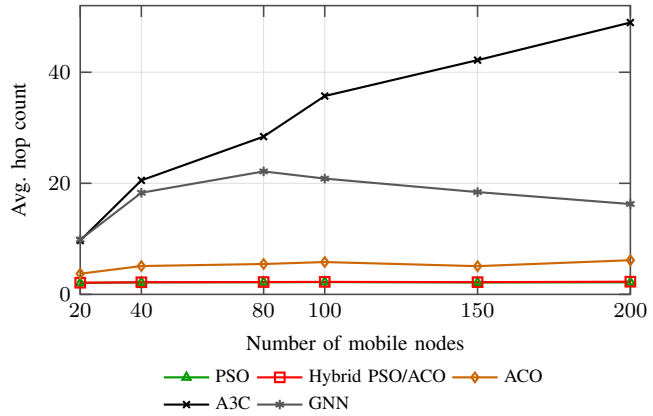


Fig. 5. Average hop count for selected algorithms. PSO and Hybrid PSO/ACO consistently find the shortest paths (2 hops), which is substantially better than A3C and GNN (9–48 hops). Shorter paths imply lower delay and higher delivery probability, explaining the advantage of swarm intelligence in latency and reliability.

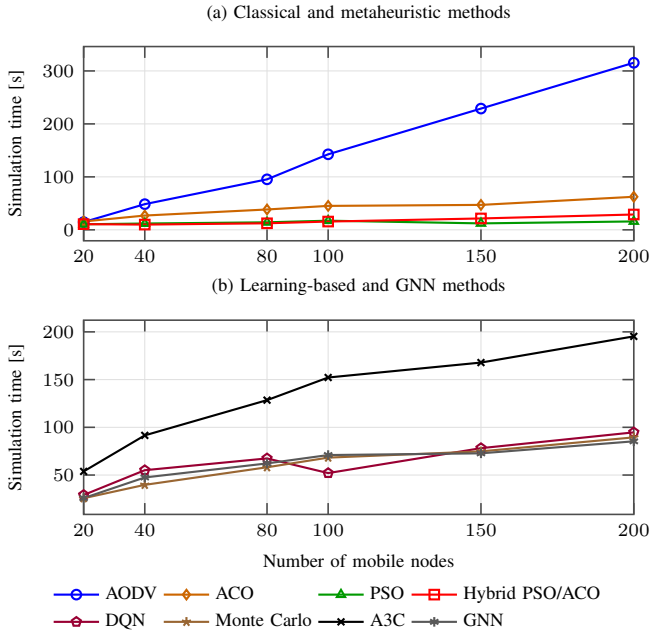


Fig. 6. Simulation runtime per algorithm. (a) Classical and metaheuristic methods: AODV increases almost linearly with node count (315 s at 200) due to growing flood-discovery overhead; ACO and PSO remain sublinear or near-constant (10–30 s), indicating good edge scalability. (b) RL and GNN: A3C is highly demanding (195 s), Monte Carlo is moderate (90 s), DQN is moderate-to-high (95 s). For practical Raspberry Pi deployment (small single-core edge processor), PSO, Hybrid PSO/ACO, ACO, and pre-trained DQN are the most realistic options.

GNN: Insufficient multi-agent coordination and expressivity without pretraining.

It is important to state the limits of the current study. Simulations used fixed mobility and traffic configurations, which may affect absolute metric values in specific deployments. Nevertheless, the relative ranking of methods is stable across scenarios and supported by multiple metrics simultaneously (PDR, delay, throughput, hop count, runtime), increasing confidence in the conclusions. This is crucial for practical edge deployment decisions, where performance, robustness, and computational cost must be balanced.

IV. Future Plans

In the next period, we plan to transfer the most promising algorithms (Hybrid PSO/ACO, DQN, ACO) to a 5G testbed with Raspberry Pi edge nodes and 5G modems, including hardware-in-the-loop validation [12]. Priorities include model transferability, energy and memory requirements, and stability under real radio interference. Extended mobility scenarios and hyperparameter sensitivity analysis will also be added to provide robust practical recommendations.

Acknowledgment

This study was partially funded by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the projects KEGA 049TUKE-4/2024, by the Slovak Research and Development Agency under the projects APVV-25-0005.

References

[1] P. Mach, Z. Becvar, and M. Draxler, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[3] R. Chataut and R. Akl, "6G networks and the AI revolution—exploring technologies, applications, and emerging challenges," *Sensors*, vol. 24, no. 6, p. 1888, 2024.

[4] J. Badár, "Artificial intelligence in routing for mobile multi-hop networks in 6G: A comprehensive review," in *Proc. 25th Scientific Conference of Young Researchers (SCYR 2025)*. Technical University of Košice, 2025, pp. 58–62. [Online]. Available: https://scyr.fei.tuke.sk/index.php/proceedings/

[5] J. Badár and J. Papaj, "Development of a custom MANET simulator with enhanced integration of AI-based routing protocols," in *Proc. 35th Int. Conf. Radioelectronics*. IEEE, 2025, pp. 1–4.

[6] J. Badár, J. Papaj, and D. Maček, "Comparison of routing algorithms in MANET using reinforcement learning methods," in *Electrical Engineering and Informatics 16—Proc. Faculty of EEI, TUKE*, 2025, pp. 560–565. [Online]. Available: https://eei.fei.tuke.sk/wp-content/uploads/2025/07/EEI_16.pdf

[7] N. Kurkina, J. Papaj, and J. Badár, "Enhancing routing efficiency in Cloud MANET using KNN and fitness function for dynamic network environments," *Peer-to-Peer Networking and Applications*, vol. 19, p. Article 40, 2026, open Access, Springer Nature.

[8] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," in *Modeling and Tools for Network Simulation*, 2010, pp. 15–34.

[9] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmüller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 529, no. 7587, pp. 529–533, 2015.

[10] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Harley, T. P. Lillicrap, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48, 2016, pp. 1928–1937.

[11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2017.

[12] L. C. Mutalemwa and S. Shin, "A classification of the enabling techniques for low latency and reliable communications in 5G and beyond: AI-enabled edge caching," *IEEE Access*, vol. 8, pp. 205 502–205 533, 2020.

Quantum Annealing in Traffic Management & Quantum–ML Synergy

[1]*Renáta RUSNÁKOVÁ (2nd year),*
Supervisor: [2]*Martin CHOVANEC*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]renata.rusnakova@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*—Quantum annealing provides a promising approach for solving large-scale combinatorial optimization problems arising in modern engineering systems. In this work, we summarize our recent research on QUBO-based formulations for traffic flow optimization and graph partitioning, with a focus on hybrid quantum–classical solution strategies using current quantum annealers. We present a data-driven framework for congestion-aware Traffic Flow Optimization and a machine learning–based method for penalty parameter tuning in the Minimum Bisection Problem. Our results demonstrate that hybrid quantum annealing can achieve near-optimal solution quality compared to classical solvers while maintaining stable runtime behavior for realistic problem sizes.**

***Keywords*—Minimum Bisection Problem, Quantum Annealing, QUBO, Optimization, Traffic Management**

I. Introduction and Motivation

Urban traffic networks and large graph-based systems exhibit complex combinatorial structures, dense interactions, and strict constraints. These characteristics make it challenging for classical optimization methods to find high-quality solutions within reasonable time limits. At the same time, recent progress in quantum hardware and hybrid quantum–classical solvers naturally raises the question of whether these approaches can already bring practical benefits to real-world optimization problems. In this context, quantum annealing has emerged as a promising approach for solving combinatorial optimization tasks. By exploiting quantum tunneling and energy-based search dynamics, it offers an alternative to standard optimization methods. A wide range of real-world engineering problems, including traffic management, clustering, and network partitioning, can be formulated as Quadratic Unconstrained Binary Optimization (QUBO) problems, which are directly compatible with quantum annealers. Therefore, our work focuses on the design of scalable, data-driven optimization frameworks that

- faithfully represent real systems,
- can be solved by current quantum annealers and hybrid solvers and
- outperform or complement classical approaches in terms of solution quality, robustness, or runtime.

II. Background and Progress

Quantum annealers, most notably those developed by D-Wave Systems, solve optimization problems expressed as Ising or QUBO models by searching for low-energy configurations of a physical Hamiltonian. This paradigm has been applied to problems, where constraints are embedded via penalty terms and interactions between decision variables are represented quadratically [1], [2]. Hybrid solvers combine quantum sampling with classical preprocessing and postprocessing to overcome hardware limits in qubit count and connectivity, enabling instances with thousands of variables to be addressed in practice [3]. While early studies demonstrated feasibility, they were typically restricted to small synthetic benchmarks or simplified networks [4], [5].

A. Initial Work on Graph Problems

In the initial phase of our research, we focused on understanding how classical graph problems can be translated into QUBO form and solved using hybrid quantum annealing. Two representative problems were investigated

1) **Shortest path problem**, with path feasibility and edge continuity constraints encoded in QUBO and the solution benchmarked against A* search.
2) **Graph partitioning problem**, where balance and cut constraints are encoded in QUBO and the solution is benchmarked against classical solvers METIS and the Kernighan–Lin algorithm.

For shortest path instances, we observed that hybrid quantum solvers can find optimal solutions for small graphs, but solution quality decreases with increasing graph size due to embedding complexity, penalty sensitivity, and noise [6]. This highlighted the central role of penalty parameter selection and model scaling. Graph partitioning proved more promising, since a carefully designed QUBO formulation and appropriately tuned constraint penalties allowed hybrid quantum annealing to compete with, and in many cases outperform, classical heuristics in terms of cut quality on mid-sized graphs.

B. Penalty Parameter Tuning as a Bottleneck

A recurring challenge across all studied problems is the selection of a penalty parameter that enforces constraints (e.g., path continuity, partition balance, one-hot assignment) without dominating the objective or producing infeasible solutions. Analytical bounds are often overly conservative, while manual tuning is infeasible for large or heterogeneous datasets. This observation directly motivated our later work on machine learning–based penalty tuning for a specialized variant of the graph partitioning problem.

III. Minimum Bisection Problem with Machine Learning–based Penalty Tuning

We addressed the Minimum Bisection Problem (MBP), a fundamental graph partitioning task with applications in parallel computing, network design, and clustering. The main difficulty is enforcing the balance constraint (equal partition sizes) while minimizing the cut size. We developed a hybrid framework that combines

- analytical bounds for penalty parameters, and
- Gradient Boosting Regression (GBR) to predict instance-specific penalty values from simple graph features (number of nodes and density).

The method was evaluated on a dataset of Erdős–Rényi graphs with $n \in \{100, \ldots, 4000\}$ nodes and edge probabilities $p \in \{0.1, 0.25, 0.5, 0.75\}$, covering a wide range of sparse to dense graphs. In total, hundreds of graph instances were generated and stored together with solver outputs for training and evaluation.

Experimental results show that penalty tuning has a significant impact on both feasibility and solution quality. The baseline analytical estimation achieved a feasibility rate of 77.93%, which increased to 100% after applying scaling and machine learning–based tuning. Using GBR-predicted parameters, the hybrid quantum solver outperformed classical heuristics (METIS) in all tested instances. The relative improvement in cut size stabilized at approximately 0.30–0.40% for larger graphs, with absolute differences reaching up to 1372 edges for $n = 4000$[7].

IV. Realistic Traffic Flow Optimization as a Large-scale QUBO Problem

The main research effort of the past year focused on formulating and solving realistic city-wide Traffic Flow Optimization (TFO) as a QUBO problem. The task is to assign exactly one route to each vehicle from a small set of alternatives such that the spatiotemporal congestion caused by vehicles sharing road segments is minimized, and individual travel times remain close to the shortest available option.

We developed a time-resolved congestion model based on sampled vehicle trajectories. Routes are discretized into temporal steps, and congestion is computed from leader–follower interactions on shared road segments in the same direction. This forms pairwise conflict costs between route alternatives, producing a dense 4-dimensional interaction tensor that captures realistic traffic behavior at high resolution [8]. This approach differs from static flow models ([4], [5]) by incorporating temporal overlap, directionality, and distance-dependent interaction strength, which are critical for representing realistic congestion phenomena.

In our QUBO formulation with analytical constraint calculation, each vehicle–route choice is represented by a binary variable $x_{i,a} \in \{0, 1\}$, where $x_{i,a} = 1$ indicates that vehicle $i \in C$ selects route alternative $a \in \{1, \ldots, k\}$. The objective consists of quadratic congestion interaction terms,

$$\sum_{i<j} \sum_{a=1}^{k} \sum_{b=1}^{k} w_{i,j,a,b}\, x_{i,a} x_{j,b}, \tag{1}$$

linear penalties reflecting route duration,

$$\sum_{i=1}^{n} \sum_{a=1}^{k} \pi_{i,a}\, x_{i,a}, \tag{2}$$

and quadratic one-hot constraints enforcing exactly one route per vehicle is selected,

$$\lambda \sum_{i=1}^{n} \Big(1 - \sum_{a=1}^{k} x_{i,a}\Big)^2. \tag{3}$$

Here, $w_{i,j,a,b} \geq 0$ denotes the congestion interaction weight between vehicle i taking route a and vehicle j taking route b, computed from spatio-temporal route overlaps, and $\pi_{i,a} \geq 0$ is the duration-related penalty for assigning vehicle i to route a. To avoid heuristic tuning, the global penalty parameter λ is derived analytically as

$$\lambda = \max_{i \in C,\, a \in \{1,\ldots,k\}} \sum_{j \in C,\, j \neq i} \sum_{b=1}^{k} w_{i,j,a,b}, \tag{4}$$

which corresponds to the maximum total interaction strength of any decision variable. With this setting, violating the one-hot constraint becomes more costly than any gain in the congestion objective, which enforces feasibility while keeping the model sensitive to congestion differences.

Direct optimization of tens of thousands of vehicles leads to QUBO matrices with millions of quadratic terms. To address this, we constructed a weighted vehicle conflict graph and applied Leiden community detection to decompose the city-scale problem into dense subproblems. Each cluster is optimized independently, and the global solution is reconstructed from the cluster assignments. This strategy preserves the dominant congestion interactions while reducing memory requirements and solver complexity by orders of magnitude.

We evaluated the framework on multiple urban road networks extracted from OpenStreetMap, with up to 25,000 vehicles and two route alternatives per vehicle. Experiments included both random origin–destination distributions and attraction-point scenarios (e.g., stadium or city center). Two solver families were compared: D-Wave's hybrid quantum annealing hybrid solver (QAHS) and exact mixed-integer optimization using Gurobi as a classical baseline. Across thousands of instances:

- the hybrid quantum solver consistently produced feasible solutions ($\geq 99.9\%$),
- the average objective gap relative to Gurobi remained below 1%, and
- runtimes were stable and dependent on instance size due to problem decomposition.

Most importantly, both solvers achieved system-wide congestion reductions of up to 25% compared with shortest-path routing, demonstrating that cooperative route assignment can significantly outperform individual greedy navigation strategies (Table I).

An additional observation from our experiments is that solver performance is strongly influenced by the structure of the underlying city graph (Fig. 1). Grid-like networks, such as Cardiff (Wales), tend to produce denser but more regular QUBO matrices, which are easier to embed on current quantum hardware. In contrast, more irregular networks, like Košice (Slovakia), result in sparser but more challenging instances from an embedding perspective. This suggests that the topology of the map itself plays an important role in quantum optimization and should be taken into account during benchmarking and practical deployment.

Overall, the results indicate that hybrid quantum annealing is no longer limited to small or illustrative examples, but can

TABLE I
RELATIVE IMPROVEMENT (%) IN CONGESTION ACHIEVED BY QAHS AND GUROBI COMPARED TO THE SHORTEST-ROUTE BASELINE; TESTED ON CITY MAP OF KOŠICE.

Vehicles (n)	QAHS vs. Shortest [%]	Gurobi vs. Shortest [%]
100	0.0	0.0
200	0.0	0.0
500	8.3	15.7
1,000	9.4	11.6
2,000	15.1	19.5
3,000	7.5	7.6
4,000	6.0	7.3
5,000	2.8	5.4
6,000	1.8	1.8
7,000	16.3	20.7
8,000	13.4	14.9
10,000	22.4	24.8
14,000	22.0	27.2
15,000	15.0	16.9
17,000	24.4	29.1
20,000	18.8	21.4
25,000	23.4	29.4

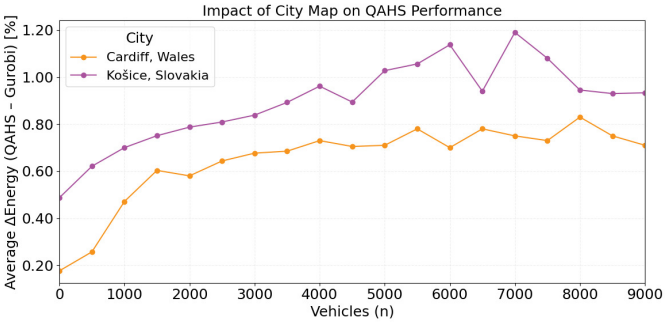


Fig. 1. Comparison of solver performance across city networks. Cardiff's uniform grid-like topology produces smaller QAHS-Gurobi energy gaps, while Košice's irregular layout results in larger deviations.

already be applied to optimization problems at a realistic urban scale, delivering competitive solution quality with consistent performance.

In summary, these results highlight three main advances achieved over the past year:

1) A general machine learning-based framework for penalty parameter tuning in QUBO models for MBP, validated on large graphs.
2) A fully data-driven, time-resolved QUBO formulation of TFO problem.
3) Demonstration that hybrid quantum annealing can match state-of-the-art classical solvers within 1% on large-scale TFO instances.

V. FUTURE DIRECTIONS

The results presented in this paper demonstrate that quantum annealing has reached a level of technical maturity that enables meaningful experimentation on realistic optimization problems. At the same time, our studies also revealed important limitations of the currently available hardware and hybrid solvers, including sensitivity to penalty parameters, embedding overhead, and limited transparency of internal optimization stages (noted also in [9]). These findings indicate that such constraints still play a key role in determining the practical applicability of quantum annealing and its hybrid optimization approaches.

Consequently, our future work will expand beyond quantum annealing and focus more broadly on gate-based quantum computing and its potential for optimization and learning tasks. In particular, we plan to investigate the Quantum Approximate Optimization Algorithm (QAOA) as a complementary paradigm for solving combinatorial problems, enabling a direct comparison between analog annealing-based approaches and variational circuit-based methods. This will allow us to assess differences in solution quality, robustness to noise, scalability, and suitability for hybrid quantum–classical workflows.

Another major research direction concerns the systematic study of noise and its algorithmic implications. While quantum annealers implicitly incorporate noise at the hardware level, gate-based platforms offer explicit control over noise models, circuit depth, and error mechanisms. We therefore intend to explore noisy quantum algorithms in controlled simulation environments, with particular attention to the Bernstein–Vazirani (BV) algorithm and its noisy variants, which are closely related to the Learning Parity with Noise (LPN) problem. This line of research is especially relevant for cryptography and post-quantum security, and directly connects quantum algorithm design with quantum hardware constraints.

From an infrastructure perspective, we plan to leverage the Qaptiva quantum simulation framework deployed within Perún HPC cluster to perform large-scale experiments with noisy quantum circuits and hybrid optimization pipelines.

Finally, an important long-term objective is education and knowledge transfer. Based on the experience gained from both quantum annealing and gate-based quantum computing research, we plan to prepare structured study materials and practical tutorials covering QUBO modeling, variational algorithms, quantum noise, hybrid optimization techniques, and quantum methods for machine learning. These resources will support students and researchers entering the field and help integrate quantum computing methods into engineering and applied informatics education.

REFERENCES

[1] G. Kochenberger, J. K. Hao, F. Glover *et al.*, "The unconstrained binary quadratic programming problem: a survey," *Journal of Combinatorial Optimization*, vol. 28, pp. 58–81, 2014.
[2] F. Glover, G. Kochenberger, R. Hennig *et al.*, "Quantum bridge analytics I: a tutorial on formulating and using qubo models," *Annals of Operations Research*, vol. 314, pp. 141–183, 2022.
[3] D-Wave Systems Inc., "Hybrid solvers for quadratic optimization," D-Wave Systems, Tech. Rep., 2022, https://www.dwavequantum.com/media/soxph512/hybrid-solvers-for-quadratic-optimization.pdf (Accessed: 2026-01-20).
[4] F. Neukart, G. Compostella, T. Seidel, D. von Dollen, S. Yarkoni, and B. Parney, "Traffic flow optimization using a quantum annealer," *Frontiers in ICT*, vol. 4, no. 29, p. 29, 2017.
[5] H. Salloum, S. Zhanalin, A. A. Badr, and Y. Kholodov, "Mini-scale traffic flow optimization: an iterative qubos approach converting from hybrid solver to pure quantum processing unit," *Scientific Reports*, vol. 15, no. 1, p. 22904, 2025.
[6] R. Rusnáková, M. Chovanec, M. Harahus, and J. Gazda, "Qubo formulation of the shortest path algorithm and its evaluation on a quantum annealer," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2025, pp. 249–254.
[7] R. Rusnáková, M. Chovanec, and J. Gazda, "Minimum bisection problem: Machine learning–based penalty parameter tuning for optimization on quantum annealers," 2026, submitted to Engineering Applications of Artificial Intelligence (Elsevier).
[8] ——, "Quantum annealing for realistic traffic flow optimization: Clustering and data-driven qubo," 2026, submitted to IEEE Access.
[9] F. A. Quinton, P. A. S. Myhr, M. Barani, P. Crespo del Granado, and H. Zhang, "Quantum annealing applications, challenges and limitations for optimisation problems compared to classical solvers," *Scientific Reports*, vol. 15, no. 1, p. 12733, 2025.

Malware Threat Detection Using Machine Learning Methods

[1]*Dušan ČATLOCH (2nd year),*
Supervisor: [2]*Eva CHOVANCOVÁ*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]dusan.catloch@tuke.sk, [2]eva.chovancova@tuke.sk

***Abstract*—Cybersecurity threats have become increasingly sophisticated, targeting information systems through various attack vectors such as malware, ransomware, distributed denial-of-service attacks, and phishing campaigns. Traditional signature-based detection techniques are often insufficient to address these threats due to their limited ability to detect novel or evolving attack patterns. As a result, machine learning and artificial intelligence techniques have gained significant attention as promising tools for improving the effectiveness and adaptability of cybersecurity systems.**

The primary motivation of this research is to investigate the applicability of machine learning algorithms for detecting and classifying cybersecurity threats in diverse scenarios. By leveraging data-driven models, it is possible to enhance detection accuracy, reduce false positives, and improve the robustness of security mechanisms against previously unseen attacks. This paper provides a summary of the research activities carried out during the past academic year, focusing on the application of machine learning techniques to cybersecurity-related problems and highlighting key results achieved through published conference contributions.

***Keywords*—Malware detection, machine learning, cybersecurity, threat recognition, anomaly detection, malware analysis.**

I. Introduction and Motivation

The increasing reliance on digital infrastructures has led to a significant rise in the frequency and complexity of cyber attacks, including malware infections, ransomware campaigns, distributed denial-of-service (DDoS) attacks, and phishing techniques. These threats pose serious risks to the availability, integrity, and confidentiality of information systems. Conventional cybersecurity solutions, which are largely based on predefined rules and signature-based detection, often struggle to adapt to rapidly evolving attack strategies and previously unseen threats.

Recent advances in machine learning (ML) and artificial intelligence (AI) have demonstrated considerable potential in addressing these limitations by enabling data-driven threat detection and classification. Machine learning models are capable of identifying complex patterns in large-scale datasets and can be continuously adapted to new attack behaviors. Motivated by these advantages, our research focuses on the systematic investigation of machine learning algorithms for cybersecurity applications, with particular emphasis on evaluating their effectiveness across multiple threat categories, including malware, ransomware, DDoS, and phishing attacks.

The primary motivation of this research is to improve the accuracy, robustness, and practical applicability of cybersecurity detection mechanisms by leveraging suitable machine learning techniques. This paper summarizes the research progress achieved during the past academic year and outlines future research directions.

II. Background and Initial Research Progress

The initial phase of the research was dedicated to building a comprehensive background in machine learning-based cybersecurity and identifying key challenges associated with threat detection in modern information systems. A review of existing detection approaches revealed that traditional security mechanisms, predominantly based on static rules and signature-based methods, have limited effectiveness against rapidly evolving cyber threats. This limitation is particularly evident in the context of malware variants, ransomware attacks, distributed denial-of-service (DDoS) campaigns, and phishing techniques, where attackers continuously modify their strategies to evade detection [1][2].

Early research activities focused on the evaluation of classical machine learning algorithms commonly applied in cybersecurity, including decision trees, random forests, support vector machines, and k-nearest neighbors classifiers. Particular emphasis was placed on data preprocessing, feature extraction, and feature selection, as cybersecurity datasets are often characterized by high dimensionality, noise, and class imbalance. The effectiveness of ensemble-based methods and the importance of proper data preparation are well documented in existing studies [3][4].

Another important challenge addressed during this stage was the detection and classification of multiple types of cyber threats within a unified framework. Studies have shown that machine learning models often suffer from limited generalization when applied across different datasets and attack categories. Variations in data sources, feature representations, and labeling practices significantly affect model performance. Furthermore, the lack of consistent datasets and standardized evaluation protocols complicates fair performance comparison and limits the transferability of trained models, highlighting the need for carefully designed experimental pipelines [5].

Overall, the initial phase of the research identified key limitations of existing cybersecurity detection techniques and confirmed the suitability of selected machine learning algorithms for threat detection tasks. The outcomes of this stage defined the direction for subsequent research activities, focusing on more comprehensive evaluations and the publication of results in international cybersecurity-oriented conferences.

III. Recent Achievements

During the assessed academic year, our research focused on the application of machine learning techniques for cybersecurity threat detection. The work addressed major categories of cyber attacks, including malware, ransomware, and distributed denial-of-service (DDoS) attacks, with the aim of evaluating the effectiveness, robustness, and practical applicability of selected algorithms. The research emphasized supervised learning approaches and systematic experimental evaluation, reflecting current trends in data-driven cybersecurity.

The studies followed a consistent methodological framework, including data preprocessing, feature selection, model training, hyperparameter optimization, and performance evaluation using relevant security metrics. Multiple machine learning models were compared within unified pipelines to ensure fair and reproducible results. The outcomes were disseminated through peer-reviewed international conference publications, forming a coherent body of work demonstrating progress in machine learning-based cybersecurity threat detection.

To improve clarity and readability, the key results of the conducted studies are summarized in Table I.

TABLE I
SUMMARY OF MACHINE LEARNING MODELS AND RESULTS ACROSS DIFFERENT STUDIES

Study	Threat Type	Models Used	Key Results
Ransomware Detection	Ransomware	RF, XGBoost, SVM	High F1-score, near 1.0 ROC-AUC
Multi-threat Detection	Malware, DDoS, Phishing	LR, KNN, DT	Accuracy ∼95%+
DDoS Detection	DDoS	RF, DT, KNN, LR	Stable performance, RF best

The following subsections summarize the key contributions and results achieved in the individual research studies carried out during the previous year, highlighting their technical contributions, experimental findings, and relevance to real-world cybersecurity applications.

A. Ransomware Detection Using Machine Learning Algorithms

One of the key research outcomes was an investigation of ransomware detection using machine learning techniques based on static analysis of Windows Portable Executable (PE) files. The study examined whether lightweight header-level features are sufficient for reliable pre-execution ransomware detection in realistic deployment scenarios.

An automated detection pipeline was implemented to extract static PE features and evaluate multiple supervised classifiers, including Logistic Regression, Support Vector Classification, Random Forest, Gradient Boosting, XGBoost, and an ensemble VotingClassifier. The experiments were conducted on a large dataset containing more than 138,000 labeled PE samples, with extensive preprocessing, feature selection, and class rebalancing applied to ensure stable and reproducible evaluation.

The results showed that tree-based and ensemble models consistently outperformed linear classifiers, achieving high F1-scores and near-ceiling ROC AUC values, while also demonstrating strong generalization to previously unseen ransomware samples. From a practical perspective, the reliance on static features enables fast and safe screening without executing binaries, making the approach suitable for deployment at internet ingress points such as mail gateways.

Overall, the study confirmed that static machine learning-based analysis can serve as an effective first line of defense against ransomware, combining high detection accuracy with scalability and low operational overhead [6].

B. Machine Learning-Based Detection and Classification of Malware, DDoS, and Phishing Attacks

A major research contribution was a comprehensive study addressing the detection and classification of multiple cybersecurity threats—malware, distributed denial-of-service (DDoS), and phishing—within a unified machine learning framework. The work focused on comparing the performance of different learning paradigms across heterogeneous datasets in order to identify their strengths and limitations in realistic cybersecurity scenarios.

The study evaluated supervised learning methods, including Logistic Regression, K-Nearest Neighbors, and Decision Trees, as well as an unsupervised clustering approach based on K-Means. Separate datasets were used for each attack category, with careful preprocessing and standardized evaluation procedures to ensure fair and reproducible comparisons.

Experimental results showed that supervised classifiers achieved high and stable performance across all threat types, with malware and phishing detection reaching accuracy levels around or above 95 % after systematic hyperparameter optimization. In contrast, unsupervised K-Means clustering consistently failed to provide meaningful classification performance despite dimensionality reduction and visual separability of clusters.

The evaluation included quantitative metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure objective performance comparison across models.

The findings clearly demonstrate that classical supervised machine learning models, when properly optimized, are effective for multi-domain cybersecurity threat detection, while unsupervised approaches exhibit significant limitations. These results emphasize the importance of labeled data and rigorous evaluation pipelines in the design of practical and reliable machine learning-based cybersecurity systems [7].

C. Machine Learning-Based Detection of DDoS Attacks in Network Traffic

Another important research outcome was a focused study on the detection of distributed denial-of-service (DDoS) attacks in network traffic using supervised machine learning techniques. The objective was to evaluate the suitability of commonly used classification models for reliable and scalable DDoS detection under realistic conditions.

The experiments were conducted on the publicly available APA-DDoS dataset using a unified pipeline that included data preprocessing, feature selection, model training, and hyperparameter optimization. To improve robustness, the preprocessing stage incorporated handling missing values, feature reduction, noise injection, and partial label flipping to better reflect real-world data imperfections.

Four supervised classifiers—Logistic Regression, K-Nearest Neighbors, Decision Tree, and Random Forest—were implemented and evaluated using identical performance metrics.

While all models achieved competitive accuracy, Random Forest consistently provided the most stable and well-generalized results across cross-validation folds.

The study highlighted important trade-offs between accuracy, interpretability, and deployment feasibility. Ensemble-based models proved particularly effective in capturing non-linear traffic patterns, confirming their suitability for real-time network monitoring and modern cybersecurity infrastructures [8].

D. Ongoing Research

In addition to the published conference contributions, further research activities were carried out during the assessed period, resulting in several manuscripts that are currently under review or in preparation. One ongoing study focuses on the integration of Security Orchestration, Automation, and Response (SOAR) and Security Information and Event Management (SIEM) systems to enhance incident detection and response capabilities in small and medium-sized organizations, examining how automated workflows and centralized event analysis can complement machine learning-based detection approaches.

Another manuscript under review examines the effectiveness of machine learning algorithms for malware detection and prediction, while a study currently in preparation extends previous research by jointly addressing malware and ransomware detection using machine learning techniques within a unified detection framework.

These ongoing activities reflect the continuation and expansion of our research agenda in machine learning-based cybersecurity and provide a foundation for future publications.

IV. Future plans

The forthcoming stage of the research will focus on advanced forms of malware, particularly polymorphic malware, which remains a significant challenge for existing detection mechanisms. Recent studies have shown that although classical and ensemble-based machine learning algorithms can achieve promising detection accuracy, their performance is often constrained by limited dataset size, class imbalance, and insufficient representation of real-world polymorphic variants. These limitations highlight the need for further validation of machine learning-based approaches on more diverse and realistic datasets [9].

In addition to polymorphic malware, future research will address metamorphic malware and advanced obfuscation techniques that significantly hinder both static and dynamic analysis. As documented in existing surveys, continuous code transformation, encryption, and API obfuscation enable malware to evade signature-based and feature-driven detection methods, posing substantial challenges to reverse engineering and reliable feature extraction [10].

To improve robustness against such evolving threats while maintaining practical deployability, the planned research will explore advanced artificial intelligence techniques, with a focus on deep learning models combined with knowledge distillation. This approach enables the transfer of knowledge from complex, high-capacity models to lightweight classifiers, allowing a balance between detection performance and computational efficiency in resource-constrained environments [11].

Future work will also consider the evaluation of these methods across heterogeneous malware families and datasets, as well as their integration into existing static and network-based detection pipelines. The ultimate objective is to develop adaptive, efficient, and deployable malware detection strategies capable of addressing emerging threats in operational cybersecurity environments.

V. Conclusion

This paper summarizes the research activities conducted during the previous academic year, focusing on the application of machine learning techniques for cybersecurity threat detection. The studies addressed multiple attack categories, including malware, ransomware, and distributed denial-of-service (DDoS) attacks, and demonstrated that classical and ensemble-based machine learning models can achieve reliable detection performance when supported by appropriate preprocessing and evaluation strategies.

The obtained results, validated through peer-reviewed conference publications, confirm the practical potential of machine learning-based approaches while also highlighting limitations related to generalization and adaptability. Future work will build on these findings by focusing on emerging malware threats and advanced artificial intelligence techniques to further improve the robustness and efficiency of cybersecurity detection systems.

References

[1] P. R. Kothamali and S. Banik, "Limitations of signature-based threat detection," 03 2022. [Online]. Available: https://www.researchgate.net/publication/388494583_Limitations_of_Signature-Based_Threat_Detection

[2] G. Apruzzese, M. Colajanni, L. Ferretti, and M. Marchetti, "Addressing adversarial attacks against security systems based on machine learning," 05 2019, pp. 1–18. [Online]. Available: https://www.researchgate.net/publication/334417231_Addressing_Adversarial_Attacks_Against_Security_Systems_Based_on_Machine_Learning

[3] V. Manjunatha and R. Ramesh, "Machine learning in malware detection: A survey of analysis techniques," *IJARCCE*, vol. 12, p. 204, 04 2023. [Online]. Available: https://www.researchgate.net/publication/378300286_Machine_Learning_in_Malware_Detection_A_Survey_of_Analysis_Techniques

[4] A. Bensaoud, J. Kalita, and M. Bensaoud, "A survey of malware detection using deep learning," *Machine Learning with Applications*, vol. 16, p. 100546, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827024000227

[5] Y. Song, D. Zhang, J. Wang, Y. Wang, Y. Wang, and P. Ding, "Application of deep learning in malware detection: a review," *Journal of Big Data*, vol. 12, 04 2025. [Online]. Available: https://www.researchgate.net/publication/391014102_Application_of_deep_learning_in_malware_detection_a_review

[6] D. Čatloch, E. Chovancová, M. Chovanec, and S. Mat'ašová, "Ransomware detection using machine learning algorithms," in *2025 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2025, pp. 133–138.

[7] D. Čatloch, E. Chovancová, M. Chovanec, and M. Štancel, "Application of machine learning algorithms for cybersecurity: Detection and classification of malware, ddos, and phishing attacks," in *2025 IEEE 29th International Conference on Intelligent Engineering Systems (INES)*, 2025, pp. 000 353–000 358.

[8] D. Čatloch, E. Chovancová, M. Chovanec, and S. Mat'ašová, "Ddos attack detection using machine learning," 2026.

[9] N. B. Sanusi, Z. I. M. Ibrahim, and I. Hassan, "Evaluation of machine learning algorithm for polymorphic malware detection," 2025.

[10] K. Brezinski and K. Ferens, "Metamorphic malware and obfuscation: A survey of techniques, variants, and generation kits," *Security and Communication Networks*, vol. 2023, no. 1, p. 8227751, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2023/8227751

[11] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International journal of computer vision*, vol. 129, no. 6, pp. 1789–1819, 2021.

Privacy-Preserving Biometric Authentication Using Advanced Cryptographic Techniques

[1]*Eva Kupcová (3rd year),*
Supervisor: [2]*Matúš Pleva*

[1,2]Department of Computer Networks, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, Slovak Republic
[1]eva.kupcova@tuke.sk, [2]matus.pleva@tuke.sk

***Abstract*—Biometric authentication systems provide an effective solution for user identification. However, the sensitive and irreversible nature of biometric data raises serious privacy and security concerns. This paper summarizes research activities carried out during the assessed period focused on the design and evaluation of privacy-preserving biometric authentication methods based on advanced cryptographic techniques. The main published outcome is a systematic and unified experimental comparison of homomorphic encryption paradigms for secure processing of biometric data, with emphasis on practical trade-offs in latency, ciphertext expansion, memory consumption, and supported computation depth. In addition, complementary work investigated robust voice-based biometrics under spoofing attacks, including speech liveness detection and morphed voice analysis, and explored secure communication mechanisms suitable for deployment in embedded and IoT environments. The reported research advances the development of secure and deployable privacy-preserving biometric authentication systems and guides future system-level integration.**

***Keywords*—biometric authentication, privacy preservation, cryptographic protection**

I. Introduction

PRIVACY-PRESERVING biometric systems have become an important research topic due to the growing deployment of biometric authentication in modern information systems, mobile platforms, and Internet of Things (IoT) environments [1]. While biometric traits offer a convenient and reliable method of user identification, their sensitive and irreversible nature introduces significant privacy and security risks. Unauthorized disclosure or misuse of biometric data may result in permanent consequences, since compromised biometric characteristics cannot be revoked or replaced in the same way as traditional credentials.

Advanced cryptographic techniques, particularly homomorphic encryption (HE), enable secure biometric processing without exposing sensitive templates [2]. Secure authentication protocols and protected communication channels are also essential in distributed and resource-constrained environments.

At the same time, biometric authentication systems must address challenges related to robustness and reliability [3]. Biometric modalities, especially voice-based systems, remain vulnerable to spoofing and presentation attacks, while contextual factors such as recording conditions and demographic variability may affect system performance. Addressing these issues is essential for deploying secure biometric authentication solutions in real-world applications.

The motivation for this paper is to study the integration of advanced cryptographic mechanisms with robust biometric processing to achieve strong privacy protection without sacrificing practical performance. This paper presents recent research results on encrypted biometric processing, resistance to spoofing attacks, and secure communication frameworks, and outlines future directions toward unified privacy-preserving biometric authentication solutions.

II. Background and Progress

The protection of biometric data has traditionally relied on template encryption and access control mechanisms. While these approaches provide a baseline level of security for stored templates, they do not fully address privacy risks arising during biometric matching, authentication, and communication.

These limitations have motivated the exploration of cryptographic techniques that enable secure processing of biometric information without exposing sensitive features in plaintext.

The early stages of the research concentrated on applying HE to biometric data protection and on integrating biometric features into secure authentication protocols. At the same time, they revealed practical challenges related to computational cost, latency, and resource consumption, particularly on constrained platforms.

Building on these observations, subsequent efforts targeted improvements in computational efficiency while also addressing robustness challenges inherent to biometric systems. The work emphasized increased resistance to spoofing and presentation attacks, which represent significant threats in real deployments, and examined how contextual factors such as emotional state, demographic variability, and acquisition conditions influence system performance.

More recent activities extended privacy-preserving biometric processing toward realistic deployment scenarios, including IoT platforms and secure communication infrastructures. This phase focused on combining biometric authentication with protected communication protocols and lightweight cryptographic mechanisms appropriate for embedded devices. By jointly addressing security requirements and deployment constraints, the research evolved from demonstrating algorithmic feasibility toward achieving system-level applicability.

III. Recent Achievements and Key Results

This section summarizes the main research achievements obtained during the assessed period. The reported results focus on strengthening the robustness, privacy, and practical deployability of biometric authentication systems, with particular

emphasis on building cryptographic foundations for secure processing. The presented research covers both algorithmic aspects of biometric security and system-level implementation in constrained as well as real-world environments.

Although the research activities addressed different parts of the problem, together they contribute to a broader privacy-preserving biometric authentication framework. In particular, the reported results span local robustness mechanisms, cryptographic protection of biometric data during processing, and secure communication and authentication protocols for distributed deployment. To provide a unifying perspective on these research directions, Fig. 1 illustrates a conceptual framework showing how these components can be combined within a privacy-preserving biometric authentication system.

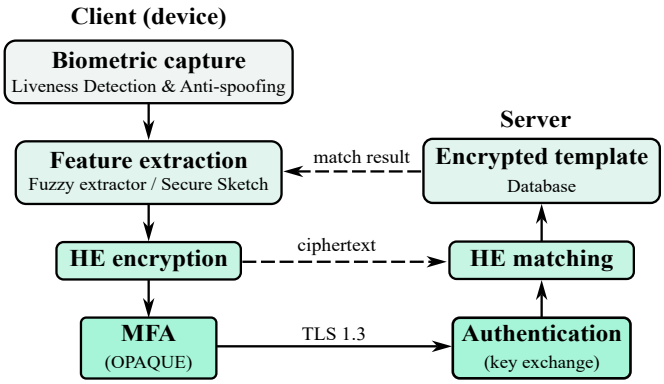


Fig. 1. Conceptual framework of a privacy-preserving biometric authentication system integrating local protection, secure transport, and encrypted matching.

A. Robust Biometric Authentication and Resistance to Spoofing Attacks

One research direction concentrated on strengthening the robustness of biometric authentication against presentation and spoofing attacks, with emphasis on voice-based modalities. Automatic speech liveness detection techniques were investigated to distinguish genuine speech from replayed or synthetically generated inputs. Experimental evaluation showed that liveness detection significantly reduces the vulnerability of voice biometric systems in attack scenarios [4].

Further research addressed morphed voice attacks in automatic speaker verification. Convolutional neural network classifiers were designed and evaluated within a real-time detection framework. Experiments conducted on a demographically diverse dataset showed that deep architectures combined with data augmentation achieve strong detection performance under clean conditions [5]. At the same time, the analysis revealed reduced accuracy for children's voices and low-quality recordings, emphasizing the need for demographic diversity and realistic training data in robust biometric systems.

In parallel, speech emotion recognition revealed how emotional content affects biometric processing. A web-based platform for automatic emotion recognition was developed to support real-time analysis and interactive experimentation. The system integrated deep learning models into an accessible interface and was evaluated on multilingual benchmark datasets [6]. The results confirmed that emotional state and cross-lingual variability significantly influence speech classification, demonstrating the need to account for contextual factors when designing reliable biometric authentication systems.

These findings extend the research toward incorporating contextual biometric attributes, including demographic and emotional information, into authentication decision processes. Ongoing efforts examine how complementary signals can strengthen primary biometric evidence in situations with degraded data quality or elevated uncertainty. Detailed methodologies and quantitative outcomes are not presented here, as this work is being prepared for separate publication.

B. Practical Evaluation of Homomorphic Encryption in Biometric Systems

The primary published research contribution of the assessed period addressed the protection of sensitive biometric data using cryptographic methods that enable secure processing without revealing biometric templates in plaintext form. A systematic experimental study compared partially, somewhat, and fully HE schemes implemented in modern cryptographic libraries under unified hardware and software conditions. The evaluation focused on deployment-relevant metrics, including encryption and decryption time, ciphertext expansion, memory consumption, and supported computation depth [7].

The results highlight practical trade-offs between functionality and computational efficiency across HE paradigms. Partially homomorphic schemes are suitable for basic aggregation tasks, whereas somewhat and fully HE enable more expressive computation at higher resource cost. The study provides criteria for choosing cryptographic protection according to application constraints, particularly in privacy-sensitive biometric systems where latency and memory are critical.

The study supports the integration of HE into biometric authentication pipelines and establishes a quantitative reference for subsequent system-level integration. It delivers a systematic and unified experimental comparison that informs when simpler homomorphic mechanisms are sufficient and when more advanced schemes become necessary for privacy-preserving biometric applications.

Concrete measurements obtained within the experimental framework illustrate the magnitude of these trade-offs. While basic homomorphic additions and decryptions are typically completed within milliseconds, multiplicative operations introduce an order-of-magnitude increase in latency. When a full refresh of computational depth is required, bootstrapping becomes the dominant factor, reaching delays on the order of seconds and expanding ciphertext sizes by several multiples. Despite these costs, the experiments confirmed stable numerical behavior and predictable resource scaling, providing input for deciding which homomorphic paradigm is technically viable for a given biometric application scenario.

C. Secure Communication Frameworks for Embedded and IoT Environments

Another research direction addressed secure communication infrastructures supporting biometric authentication, with emphasis on embedded and resource-constrained devices. An experimental platform for client–server communication based on TLS 1.3 was developed to enable the integration and evaluation of post-quantum cryptographic algorithms [8]. The platform supports systematic testing of key encapsulation mechanisms and digital signature schemes within the TLS protocol and allows assessment of their performance and deployment feasibility in post-quantum scenarios.

In parallel, a lightweight TLS 1.3 client was implemented for the ESP32 microcontroller. The design relies on a reduced set of cryptographic primitives to maintain interoperability with standard web servers while limiting memory footprint and computational overhead. Experimental results confirmed that secure TLS communication is achievable on embedded hardware without unacceptable performance penalties [9].

Beyond basic connectivity, the developed infrastructure enables systematic experimentation with alternative cryptographic primitives and certificate configurations. The framework has been used to evaluate a wide spectrum of key encapsulation and digital signature mechanisms, allowing comparative analysis of computational demands across different hardware platforms. The measurements highlighted substantial diversity in performance characteristics among candidate algorithms and confirmed that implementation choices have a decisive impact on feasibility in constrained environments. These insights provide an important foundation for subsequent optimization and for the gradual transition toward quantum-resistant secure communication.

Additional ongoing work examines implementation aspects of password-authenticated key exchange protocols relevant to embedded environments, including OPAQUE. The investigated directions emphasize constant-time execution, secure memory management, and modular architectural design suitable for microcontroller-class platforms. Detailed protocol parameters and quantitative results are not reported here due to the ongoing review and publication process.

D. Multifactor Client–Server Authentication over TLS 1.3

During the assessed period, research focused on the design of a multifactor client–server authentication framework that integrates cryptographic credentials, device binding, and biometric verification within a TLS 1.3 secured channel. TLS 1.3 provides confidentiality, integrity, and server authentication at the transport layer, while application-layer mechanisms strengthen assurance of identity and device authenticity.

The protocol follows a challenge–response model. After password verification, the server generates a fresh random challenge and delivers it to the client over an established TLS session. The challenge is processed by a dedicated on-device module that enforces local biometric verification. Only after successful user verification does the module produce a cryptographic signature over a structured payload containing the challenge and additional control information.

The signed payload binds the response to a specific session and service context and incorporates a monotonic signature counter associated with the device and user account. The server validates the counter against its stored state, enabling detection of replay attempts, unauthorized signature creation, or potential key cloning. Authentication is rejected whenever the observed counter deviates from the expected progression.

To reinforce the link to the physical device, the payload may also include an identifier associated with the enrolled biometric state maintained locally. This value does not reveal biometric information and serves exclusively for integrity and audit purposes, confirming that the signature originated from the registered biometric context.

By combining server-generated challenges, device-bound keys, monotonic counters, contextual binding, and enforced on-device biometric verification, the server obtains strong assurance that authentication was executed by the intended module on the legitimate device. Detailed protocol parameters and performance characteristics are not reported here, as the system remains under active development.

IV. Conclusion and Future Research Directions

This paper summarizes research activities aimed at developing privacy-preserving biometric systems supported by advanced cryptographic techniques. The results show that cryptographic protection can be combined with biometric authentication to achieve secure processing, improved resistance to spoofing, and protected communication in practical deployments. The presented work confirms the viability of integrating HE, secure authentication protocols, and robustness-enhancing mechanisms within a unified framework.

Future research will focus on integrating the evaluated components into a comprehensive biometric authentication application. The plan includes combining cancelable biometrics, fuzzy extractors, and secure sketch constructions with cryptographic protocols such as OPAQUE and TLS 1.3 within a shared experimental environment.

Further efforts will target system-level optimization and validation in resource-constrained settings, particularly for IoT and embedded devices, while considering post-quantum cryptographic techniques for long-term security. Emphasis will remain on experimental validation using real sensors and representative datasets, with attention to robustness, scalability, and deployability.

Acknowledgment

The research was partially supported by the Ministry of Education, Research, Development, and Youth of the Slovak Republic under the project KEGA 049TUKE-4/2024 & VEGA 1/0685/26, by the Slovak Research and Development Agency under the projects APVV-22-0414 & APVV-22-0261 as well as by the Erasmus+ MetaCog 2024-1-FI01-KA220-HED-000247530 project.

References

[1] K. K. Prakasha and U. Sumalatha, "Privacy-preserving techniques in biometric systems: Approaches and challenges," *IEEE Access*, 2025.

[2] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys (Csur)*, vol. 51, no. 4, pp. 1–35, 2018.

[3] C.-W. Lien and S. Vhaduri, "Challenges and opportunities of biometric user authentication in the age of IoT: A survey," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–37, 2023.

[4] R. Haluška, E. Kupcová, M. Pleva, and F. Hric, "Automatic speech liveness detection," in *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 2024, pp. 168–173.

[5] E. Kupcová, M. Pleva, Z. Sokolová, and M. Novotný, "Real-time detection of morphed voices in biometric speaker verification systems," in *2025 IEEE 25th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2025, pp. 155–160.

[6] E. Kupcová, M. Pleva, Z. Sokolová, K. Sopkovič, M. Lukáč, and D. Levický, "Web-based implementation of automatic speech emotion recognition," in *2025 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2025, pp. 500–505.

[7] E. Kupcova, M. Pleva, V. Khavan, and M. Drutarovsky, "A comparative study of partially, somewhat, and fully homomorphic encryption in modern cryptographic libraries," *Electronics*, vol. 14, no. 23, 2025. [Online]. Available: https://doi.org/10.3390/electronics14234753

[8] E. Kupcova, J. Simko, M. Pleva, and M. Drutarovsky, "Experimental framework for secure post-quantum TLS client-server communication," in *2024 International Symposium ELMAR*, 2024, pp. 213–216.

[9] J. Pajkos, E. Kupcova, M. Pleva, and M. Drutarovsky, "Esp32 microcontroller based lightweight TLS 1.3 client for IoT applications," in *2025 35th International Conference Radioelektronika (RADIOELEKTRONIKA)*, 2025, pp. 1–6.

Reconstructive Approach for Wearable Blood Pressure Monitoring

[1]*Jerguš ŠEVEC (1st year)*
Supervisor: [2]Ondrej KOVÁČ

[1,2]Dept. of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]jergus.sevec@tuke.sk, [2]ondrej.kovac@tuke.sk

***Abstract*—Bioimpedance measurement methods for hemodynamic parameter extraction and monitoring constitute a promising alternative to currently employed techniques involving bulky cuffs. The use of this technology in wearable devices could significantly improve possibilities for cardiovascular disease monitoring, treatment, and prevention. Conventional signal acquisition adheres to the Nyquist sampling principle, which results in increased demands on analogue-to-digital converters. In recent years, compressed sensing has emerged as an alternative approach to signal acquisition that enables sub-Nyquist sampling under sparsity assumptions. The integration of reconstruction based on compressed sensing algorithms, therefore, offers the potential to reduce energy consumption and significantly prolong battery life in wearable bioimpedance systems.**

***Keywords*—Compressed sensing, bioimpedance, hemodynamic parameter estimation**

I. INTRODUCTION

Cardiovascular diseases (CVD) represent a leading cause of global mortality, causing one in three deaths worldwide [1]. Most of the recorded instances can be associated with a heart attack or stroke. Statistics indicate a direct connection between patients suffering from CVD and one or more risk factors, such as hypertension, smoking, or other factors, with blood pressure (BP) being a dominant marker [2]. Reliable and continuous monitoring of BP and related hemodynamic parameters could therefore significantly improve early diagnosis, treatment selection, and preventive strategies, while also reducing the socioeconomic burden of CVD.

The conventional measurement of systolic and diastolic blood pressure relies on cuff-based sphygmomanometry, which is highly impractical for continuous or long-term monitoring. Several studies have been developed to assess the possibilities of compact cuff-less hemodynamic indicator monitoring in patients using electrocardiogram (ECG), photoplethysmography (PPG), and bioimpedance approaches. Although the physical approach to measurement differs, these generally rely on the acquisition of pulse transition time (PTT) and pulse wave velocity (PWV), BP influences pulse wave velocity through pressure-dependent arterial wall mechanics. However, the relationship between PTT and blood pressure is nonlinear, subject-specific, and modulated by vascular tone and pre-ejection effects [3]. PPG typically involves measurements at the fingertip [4], which limits usability during motion and long-term monitoring. It is often combined with ECG [5] to estimate PTT and PWV parameters, requiring multiple sensors placed at different body locations, increasing system complexity and reducing practicality for wearable use.

Bioimpedance measurements are more practical in this respect, since these parameters can be obtained from two points near the wrist artery, allowing for the use of a single sensor patch with integrated electrodes, however they remain sensitive to electrode placement and motion artefacts [6]. Although ioimpedance systems enable compact design, conventional implementations typically employ either fixed-frequency sine, swept-sine, or square-wave excitation for spectral analysis and rely on the Nyquist theorem [7], [8]. Such approaches can lead to higher sampling rates and increased energy consumption, mainly due to analogue-to-digital converter (ADC) requirements. To overcome these limitations, a new approach to signal acquisition from a subsampled dataset has been proposed in [9], [10]. By lowering the demands on ADCs, combined with a broadband excitation signal, the energy demands of such applications could be lowered while extending associated longevity, although reconstruction accuracy depends on signal sparsity and algorithmic stability. However, the integration of compressed sensing–based system identification with wearable bioimpedance pulse wave measurements remains relatively unexplored. In this paper, the possibility of the implementation of such an approach is explored.

II. GENERAL THEORY OF COMPRESSED SENSING

The methodology of signal reconstruction from subsampled data applies to signals that are considered sparse, meaning highly compressible, in a certain domain, such as those of the Fourier, wavelet, or discrete cosine transform. A signal is considered sparse if most of the information is concentrated among k non-zero coefficients, while this number is at the same time significantly smaller than the total number of elements of a certain basis L, such that $k \ll L$.

A. The Fundamental principles

The fundamental idea involves a standard measurement model (1), where $y \in R^m$ denotes the vector of measurements, $x \in R^n$ represents a signal that is being reconstructed, $A \in R^{m \times n}$ corresponds to the measurement matrix and $w \in R^n$ models noise.

$$y = Ax + w \tag{1}$$

The measurement matrix A is formed by two separate matrices, the sensing matrix $\Phi \in R^{m \times n}$ and the sparsifying

matrix $\Psi \in R^{n\times n}$, as $A = \Phi\Psi$. For the reconstruction to be possible, both the sensing matrix and the sparsifying basis need to be incoherent. This means that the rows of one matrix must be independent of the columns in the other. The mutual coherence is denoted as C and directly influences the approximate minimal number of measurements M required for stable reconstruction, as seen in (2) [11].

$$M \geq \left(Ck \log_{10} \frac{L}{k}\right) \tag{2}$$

For the reasons of incoherence, random matrices are utilised for sensing purposes. The most widely used types are based on the Gaussian and Bernoulli distributions [12]. These are universally incoherent and provide good results as it comes to the low amount of measurements. The major drawback lies in their unstructured nature, requiring significant storage for data and being slow in processing. This limitation is commonly solved by utilising random structured matrices, whose components can be generated according to a predetermined algorithm, thus reducing the needed storage and speeding up the computations.

If both the conditions of minimum number of measurements M and the low enough coherence between the sensing and transform matrices are satisfied, then the system approximately preserves the Euclidean distances between components. This property is commonly referred to as the Restricted Isometry Property and is a crucial condition for successful reconstruction. In many cases, this property is generally difficult to determine [10], leaving the mutual coherence as the main indicating parameter.

B. The Reconstruction Dilemma

The undersampling inherent to the sensing process creates an underdetermined system. Since the number of measurements M is significantly smaller than the signal length to be reconstructed N, there are infinitely many possible combinations of vector x that allow for reconstruction through (1). This dilemma is solved by algorithms that iteratively minimise a chosen norm [12]. In theory, one could enforce sparsity directly through the amount of non-zero coefficients as in the L_0 norm, but this approach represents a non-convex problem, meaning there is no guarantee of a global optimum existence, while at the same time it is computationally demanding. In common practice, other norms such as the L_1 norm in relaxation algorithms [12], representing a total sum of absolute values of non-zero elements, or the L_2 norm, defined as the Euclidean norm of the residual vector, are used. Although relaxation algorithms, such as those of the Basis Pursuit developed in the early stages of compressed sensing theory [13], simplify the problem by finding a global optimum through convex algorithms, as in the CVX toolbox, and are guaranteed to converge to a proper solution, they require significantly more processing power than greedy algorithms and aren't optimised for time-critical implementations. On the other hand, greedy algorithms, such as the Orthogonal Matching Pursuit (OMP) or Iterative Hard Thresholding (IHT) and their derivatives, are based on faster signal decomposition through least squares principles. This decomposition does not force sparsity directly; therefore, other mechanisms are employed to do so.

The OMP algorithm described in [14] represents the most frequently used greedy algorithm [12]. It starts by initialising the residual and identifying the most correlated column of the sensing matrix with the residual vector. The most correlated coefficient is then added to the support vector, and the least-squares operation is performed. The residual vector is then updated as the difference between the measurement vector y and the transformed vector $A\hat{x}$. In this fashion, the algorithm updates the estimate until a predetermined number of coefficients k is extracted, or the norm falls below a set threshold.

The IHT algorithm [15] follows an iterative thresholding approach. It updates the estimate by taking a gradient descent step to minimise the residual L_2 norm, followed by a hard thresholding operation that retains only the k largest coefficients to enforce sparsity in each iteration. During the hard-thresholding step, the vector is sorted in descending order, while at the same time, the component indexes are preserved.

III. SYSTEM RESPONSE IDENTIFICATION

In recent years, a number of literature was published on the possibilities of compressed sensing implementation in system identification [16], [17], [18], [19]. The possibility of monitoring the physical layer of wireline channels was explored in [16]. The authors demonstrated the possibility of implementation of such a monitoring approach in transceiver front-ends. Further literature was focused in measurement of transfer characteristics of digital-to-analogue converters [17]. Another implementation was explored for early-warning monitoring of direct current motors through monitoring of impedance changes in coil windings [18], [19].

A. Methodology of Implementation

The approach explored in the mentioned studies generally combines a wide-band excitation signal with a fixed sampling frequency at the output of the considered system under test. The sampling frequency is lower than the equivalent Nyquist frequency given by the compression ratio as $CR = f_{NYQ}/f_{ADC}$. The system under test is excited using a pseudorandom binary sequence (PRBS) of ± 1 symbols with frequency set according to the Nyquist theorem. This sequence can be expressed mathematically as:

$$x(t) = 2A \sum_{m=-L}^{N-1} \left(x(m)p\left(t - mT_p\right) - \frac{1}{2}\right) \tag{3}$$

The parameter $p(t)$ corresponds to a square signal of period $T_p \in R$ as denoted in (4).

$$p(t) = \begin{cases} 1, t \in \langle 0, T_p \rangle \\ 0, otherwise \end{cases} \tag{4}$$

The system output $y(t)$ is a product of a convolution operation between the excitation signal $x(t)$ and the system impulse response $h(t)$ as well as by the noise w(t) with various origins.

$$y(t) = x(t) * h(t) + w(t) \tag{5}$$

This behaviour can be further linearised to the standard linear measurement model (1) as expressed in (6), where the measurement matrix A described earlier consists of the Toeplitz matrix $X \in R^{m\times n}$ constructed from the PRBS sequence as denoted in (7) and the diagonal matrix $R \in R^{m\times n}$ as $A = RX$. The decimation process is modelled by a diagonal

vector consisting of an equidistant series of ones and zeros corresponding to the implemented compression ratio CR. By changing the compression ratio, a different combination of linear equations is extracted as the reconstructed impulse response is inherently dispersed throughout the entire duration of PRBS.

$$y = RXh + w \tag{6}$$

Another complication concerns the use of Toeplitz matrices that naturally arise from system identification theory. Generally, they exhibit high coherence due to their internal structure composed of a timeshifted sequence. Literature shows that, given good autocorrelation properties of the construction signal, such as the suitable autocorrelation properties of PRBS, the coherence stays sufficiently low.

$$X = \begin{bmatrix} x(0) & x(-1) & \cdots & x(-L+1) \\ x(1) & x(0) & \cdots & x(-L) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & \cdots & x(N-L) \end{bmatrix} \tag{7}$$

B. On the possibilities of hemodynamic parameter estimation

The human wrist consists of multiple tissue layers with different electrical properties, forming a volume conductor with time-varying impedance [20]. During each cardiac cycle, arterial blood volume changes, which modulate the local electrical conductivity and result in a measurable variation of bioimpedance superimposed on the static impedance of surrounding tissues. By injecting a small, safe excitation current and measuring the resulting voltage, the time-varying impedance signal can be acquired. From a system identification perspective, the bioimpedance measurement can be described as the response of a quasi-linear, time-varying system, whose effective parameters are modulated by the propagating arterial pulse wave. Consequently, the pulsatile blood flow induces a dynamic system response that carries hemodynamic information.

This behaviour can be modelled as a three-dimensional structure of basic elements, voxels [21], each represented by a parallel combination of capacitance and resistance, as shown in Fig. 1. The resistive component R_T models ionic conduction in intracellular and extracellular fluids [7], while the capacitive component C_T represents the cell membrane polarization effects.

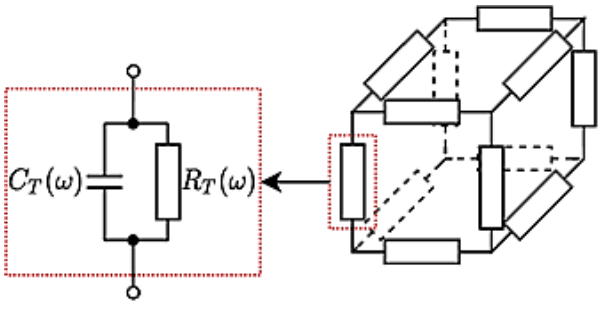


Fig. 1 Equivalent model of the cellular structure of the human wrist with a model of its elemental part [21]

In practical implementations of a single-site measurement, a multi-electrode topology is required. As illustrated in Fig. 2, the electrodes are arranged in a longitudinal array along the radial artery. This setup employs a tetrapolar sensing method to minimise the influence of skin-electrode interface impedance. By spatially separating the excitation and sensing electrodes, the measured signal predominantly reflects impedance variations originating from deeper tissue layers, which are associated with pulsatile blood flow.

The authors in [21] simulated the system identification methodology by utilising a four-lead topology (single measurement site) and successfully measured the change in volume of the flowing blood associated with the pulse using the OMP algorithm and exploited the sparsity in the time domain. To obtain the PTT parameter, a second measurement site is needed. This can be achieved by introducing an additional set of sensing electrodes between the excitation electrodes, forming an octopolar configuration in a single electrode array. Based on electrical tissue models established in literature [21], the impulse response is assumed to exhibit compressibility in the time domain, which is further exploited for signal reconstruction using PRBS excitation. This property arises because the energy of the stable system's impulse response is concentrated within a limited initial time period, allowing for an effective k-sparse approximation.

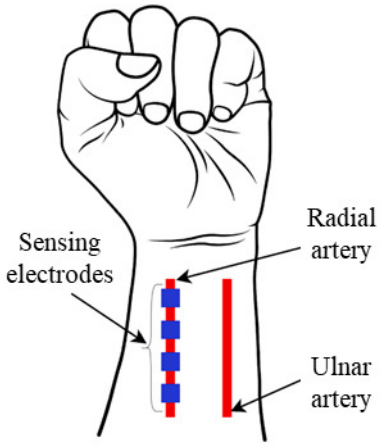


Fig. 2 Multi-electrode placement strategy for a single-site bioimpedance measurement on the human wrist [21]

The PTT parameter is defined as the time required for the pulse wave to propagate between two measurement sites. In common practice, the pulse onset is detected using the tangent method [6], where the derivative of the waveform is computed to locate the point of maximum slope (MS) on the rising edge (see Fig. 3). A tangent line is then constructed at this point, and the pulse arrival time is determined from the intersection of this tangent with the baseline. The PTT is obtained as the time difference between the pulse onsets detected in both waveforms. The PWV is subsequently derived as the ratio of the distance between the measurement sites to the corresponding PTT.

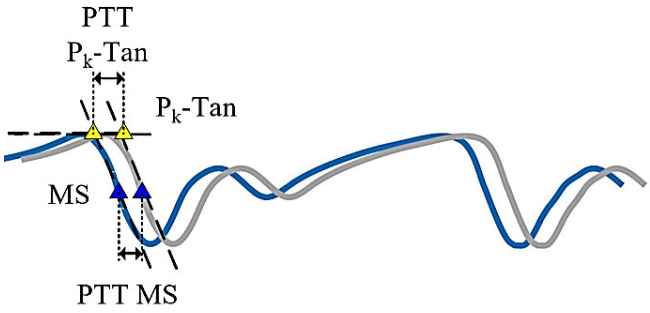


Fig. 3 Graphic representation of PTT measurement based on two bioimpedance measurement sites [6]

Practical employment of wearable bioimpedance systems usually introduces motion artefacts, which predominantly affect waveform morphology and temporal features, while

often being accompanied by amplitude variations. Signal normalisation is applied solely to map the reconstructed waveforms into a fixed numerical range [6], facilitating consistent processing and visual comparison. No explicit compensation for noise or motion-induced artefacts is performed at this stage. In contrast, motion artefacts significantly affect the rising edge morphology, particularly the MS region exploited by the tangent method [22]. Since the discussed reconstructive approach relies on preserving temporal features of the system response rather than absolute magnitude, distortions of the MS region pose a critical challenge. Therefore, robustness to motion-induced morphological distortion is a key requirement for accurate PTT estimation in bioimpedance measurements using compressed sensing.

IV. ON PROSPECTS OF FUTURE WORK

The bioimpedance measurement approach to pulse wave estimation seems to be a promising method for CVD prevention. The literature indicates the possibility of blood pressure estimation based on hemodynamic parameters with several advantages over conventional PPG [20] or its combination with ECG. PPG suffers from low skin penetration, limiting its sensitivity to changes in outer skin layers. In contrast, the excitation signal used in bioimpedance measurement propagates into deeper tissue layers and is sensitive to capacitive and resistive changes. The broader scope of application includes the acquisition of respiration, hydration or other tissue-related diagnostic parameters. The use of a single electrode patch constrains the device to a single site, unlike multisensory setups requiring synchronised ECG measurements. The overview outlines the principles of reconstructive sub-sampled signal acquisition and the conditions required for successful signal recovery. Furthermore, the methodology explores state-of-the-art approaches in system identification and impulse response reconstruction based on compressed sampling principles. The paper also discusses the feasibility of integrating these methods into pulse wave measurement systems, highlighting their potential for wearable applications. Future research will focus on developing a hardware prototype for real-time hemodynamic parameter acquisition using compressed sensing–based reconstruction. However, several challenges remain. Unlike PPG, which is highly localised and typically exhibits a higher signal-to-noise ratio, bioimpedance measurements have a higher noise floor and are prone to motion artefacts. Electrode placement relative to the radial artery is also critical, as misplacement attenuates measurable impedance differences [6]. A comparative analysis of compressed sensing algorithms will be essential to identify an optimal trade-off between energy consumption, processing speed and complexity.

ACKNOWLEDGEMENT

The work is part of a project supported by the Cultural and Educational Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic (No. 025TUKE-4/2026).

REFERENCES

[1] B. A. Stark *et al.*, “Global, Regional, and National Burden of Cardiovascular Diseases and Risk Factors in 204 Countries and Territories, 1990-2023”, *JACC*, roč. 86, č. 22, s. 2167–2243, dec. 2025, doi: 10.1016/j.jacc.2025.08.015.

[2] F. D. Fuchs a P. K. Whelton, “High Blood Pressure and Cardiovascular Disease”, *Hypertension*, roč. 75, č. 2, s. 285–292, feb. 2020, doi: 10.1161/HYPERTENSIONAHA.119.14240.

[3] Y. Ma *et al.*, “Relation between blood pressure and pulse wave velocity for human arteries”, *Proc. Natl. Acad. Sci.*, roč. 115, č. 44, s. 11144–11149, okt. 2018, doi: 10.1073/pnas.1814392115.

[4] G. Chen, L. Zou, a Z. Ji, “A review: Blood pressure monitoring based on PPG and circadian rhythm”, *APL Bioeng.*, roč. 8, č. 3, s. 031501, sep. 2024, doi: 10.1063/5.0206980.

[5] M. Y. Ho, H. M. Pham, A. Saeed, a D. Ma, “WF-PPG: A Wrist-finger Dual-Channel Dataset for Studying the Impact of Contact Pressure on PPG Morphology”, *Sci. Data*, roč. 12, č. 1, s. 200, feb. 2025, doi: 10.1038/s41597-025-04453-7.

[6] B. Ibrahim, D. Mrugala, a R. Jafari, “Effects of Bio-Impedance Sensor Placement Relative to the Arterial Sites for Capturing Hemodynamic Parameters”, v *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany: IEEE, júl. 2019, s. 6569–6573. doi: 10.1109/EMBC.2019.8857585.

[7] Y. Xin, “Adopting a Better Bioimpedance Analog Front End for Next-Generation Vital Signs Monitoring Devices”, č. 53, mar. 2024, [Online]. Available at: https://www.analog.com/en/resources/analog-dialogue/articles/adopting-a-better-bioimpedance-afe.html

[8] J. Xu *et al.*, “Analog Front-End Circuit Techniques for Wearable ExG, BioZ, and PPG Signal Acquisition: A Review”, *IEEE Open J. Solid-State Circuits Soc.*, roč. 5, s. 251–268, 2025, doi: 10.1109/OJSSCS.2025.3610583.

[9] D. L. Donoho, “Compressed sensing”, *IEEE Trans. Inf. Theory*, roč. 52, č. 4, s. 1289–1306, apr. 2006, doi: 10.1109/TIT.2006.871582.

[10] E. J. Candes a M. B. Wakin, “An Introduction To Compressive Sampling”, *IEEE Signal Process. Mag.*, roč. 25, č. 2, s. 21–30, mar. 2008, doi: 10.1109/MSP.2007.914731.

[11] S. L. Brunton a J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, 1st vyd. Cambridge University Press, 2019. doi: 10.1017/9781108380690.

[12] S. Hosny, M. W. El-Kharashi, a A. T. Abdel-Hamid, “Survey on compressed sensing over the past two decades”, *Mem. - Mater. Devices Circuits Syst.*, roč. 4, s. 100060, júl. 2023, doi: 10.1016/j.memori.2023.100060.

[13] S. S. Chen, D. L. Donoho, a M. A. Saunders, “Atomic Decomposition by Basis Pursuit”, *SIAM J. Sci. Comput.*, roč. 20, č. 1, s. 33–61, jan. 1998, doi: 10.1137/S1064827596304010.

[14] J. A. Tropp a A. C. Gilbert, “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit”, *IEEE Trans. Inf. Theory*, roč. 53, č. 12, s. 4655–4666, dec. 2007, doi: 10.1109/TIT.2007.909108.

[15] T. Blumensath a M. E. Davies, “Iterative hard thresholding for compressed sensing”, *Appl. Comput. Harmon. Anal.*, roč. 27, č. 3, s. 265–274, nov. 2009, doi: 10.1016/j.acha.2009.04.002.

[16] L. De Vito, F. Picariello, S. Rapuano, I. Tudosa, a L. Barford, “A Compressive Sampling-Based Channel Estimation Method for Network Visibility Instrumentation”, *IEEE Trans. Instrum. Meas.*, roč. 69, č. 5, s. 2335–2344, máj. 2020, doi: 10.1109/TIM.2019.2947986.

[17] L. De Vito, F. Picariello, S. Rapuano, a I. Tudosa, “A new built-in loopback test method for Digital-to-Analog Converter frequency response characterization”, *Measurement*, roč. 182, s. 109712, sep. 2021, doi: 10.1016/j.measurement.2021.109712.

[18] J. Kromka *et al.*, “Online impedance estimation of induction motor coils using a CS-based measurement method”, *Meas. Sens.*, roč. 38, s. 101425, máj. 2025, doi: 10.1016/j.measen.2024.101425.

[19] O. Kovac *et al.*, “A CS-based measurement method for impedance estimation of induction motor coil”, *Measurement*, roč. 259, s. 119586, feb. 2026, doi: 10.1016/j.measurement.2025.119586.

[20] B. Ibrahim, D. A. Hall, a R. Jafari, “Pulse Wave Modeling Using Bio-Impedance Simulation Platform Based on a 3D Time-Varying Circuit Model”, *IEEE Trans. Biomed. Circuits Syst.*, roč. 15, č. 1, s. 143–158, feb. 2021, doi: 10.1109/TBCAS.2021.3059211.

[21] J. Kromka, J. Saliga, O. Kovac, L. De Vito, F. Picariello, a I. Tudosa, “Radial artery pulse wave estimation by compressed sensing measurements of wrist bio-impedance”, *Measurement*, roč. 219, s. 113174, sep. 2023, doi: 10.1016/j.measurement.2023.113174.

[22] J. Y. A. Foo, S. J. Wilson, G. R. Williams, M. Harris, a D. M. Cooper, “Motion artefact reduction of the photoplethysmographic signal in pulse transit time measurement”, *Australas. Phys. Eng. Sci. Med.*, roč. 27, č. 4, s. 165–173, dec. 2004, doi: 10.1007/BF03178645.

Scene Understanding and Corner-Case Handling in Autonomous Driving: A Review

[1]*Daniel ŠOLC (1st year),*
Supervisor: [2]*Eva CHOVANCOVÁ*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]daniel.solc@tuke.sk, [2]eva.chovancova@tuke.sk

***Abstract*—Autonomous driving systems must perceive complex traffic environments with high reliability, yet standard perception pipelines remain vulnerable to corner cases — rare, safety-critical scenarios that deviate from training distributions. This review surveys research at the intersection of scene understanding and corner-case handling in autonomous driving, covering object detection, semantic segmentation, Bird's Eye View perception, 3D occupancy prediction, scene graph generation, open-vocabulary methods, and world-model-based simulation. Four critical open problems are identified: reliable detection of novel objects under real-time constraints, scene graph construction for open-world scenes, scalable and semantically diverse corner-case generation, and the absence of a standardized evaluation methodology for long-tail perception. A research agenda is proposed comprising open-vocabulary scene graph generation, world-model-driven corner-case synthesis, and reasoning-augmented hazard interpretation — three contributions designed to make autonomous driving perception robust in the diverse and unpredictable conditions of real-world operation.**

***Keywords*—Autonomous driving, corner cases, open-vocabulary detection, scene graphs, scene understanding.**

I. Introduction

Autonomous driving (AD) has progressed rapidly in recent years, driven by advances in deep learning, sensor technology, and large-scale datasets [1]. The perception module, responsible for interpreting the surrounding environment, is a critical component of any AD system. It must reliably detect objects, understand road layouts, and reason about spatial relationships to enable safe planning and control [2]. Despite substantial improvements, perception systems still exhibit brittleness when confronted with scenarios that lie outside their training distributions, commonly referred to as corner cases [3].

Corner cases encompass a broad range of situations: unusual objects on the road (fallen cargo, animals), rare weather or lighting conditions, atypical road geometries, and unexpected behaviors by other traffic participants [4]. These events follow a long-tail distribution, meaning they are individually rare but collectively responsible for a significant share of potential failures [5]. Addressing them is essential for achieving the level of reliability required for higher levels of driving automation.

This paper reviews the current state of research in scene understanding and corner-case handling for autonomous driving, critically examines existing approaches, and identifies open problems and promising directions for future work.

II. Literature Review and Analysis

A. Perception Foundations for Autonomous Driving

Modern AD perception systems rely on deep learning methods for object detection, semantic segmentation, and depth estimation, typically operating on data from cameras, LiDAR, and radar [6]. Convolutional neural networks (CNNs) and, more recently, vision transformers form the backbone of most architectures. Object detection approaches have evolved from two-stage detectors such as Faster R-CNN to efficient single-stage methods and transformer-based architectures such as DETR and its variants. Zhao et al. [1] provide a comprehensive survey covering both layered and end-to-end system architectures, evaluating sensor technologies for object detection and semantic segmentation. They note that while modular approaches offer interpretability and easier debugging, they suffer from information loss and error accumulation at module boundaries. Contreras et al. [7] specifically review 3D object detection approaches for real-time AD, comparing methods based on monocular vision, stereo cameras, LiDAR point clouds, and multi-modal fusion. They highlight that sensor fusion remains essential for robust detection in dynamic urban settings where occlusion, varying lighting, and adverse weather conditions are common, and that camera-LiDAR fusion methods consistently outperform single-modality approaches in challenging scenarios.

Feng et al. [6] analyze multi-modal fusion for detection and segmentation, finding that sensor-specific failure modes (camera degradation in heavy rain, LiDAR reflections on wet surfaces) propagate through the fusion pipeline, motivating more robust scene representations.

A significant trend is the shift toward end-to-end frameworks that jointly optimize perception, prediction, and planning within a single differentiable model. UniAD [8] demonstrated that a unified query-based design achieves strong performance across tracking, mapping, motion forecasting, occupancy prediction, and planning, avoiding the information bottleneck of modular pipelines. Chen et al. [2] survey end-to-end AD methods and identify long-tail scenario handling as the most pressing open problem, noting that end-to-end systems inherit the distributional biases of their training data and perform poorly on underrepresented scenarios.

B. Bird's Eye View Perception and 3D Occupancy Prediction

A significant shift in autonomous driving perception has been the move toward Bird's Eye View (BEV) representations,

which project multi-camera or LiDAR inputs into a unified top-down feature space. BEV-based methods such as BEVFusion [9] and BEVFormer [10] enable consistent multi-modal feature alignment and have become a dominant paradigm for tasks including 3D object detection, lane segmentation, and map prediction. By aggregating spatial information into a canonical coordinate frame, BEV methods avoid the perspective distortions and occlusion ambiguities that impair camera-based perception in complex urban scenarios.

Complementing BEV detection, 3D occupancy prediction methods such as TPVFormer [11] and SurroundOcc [12] predict fine-grained volumetric occupancy of the scene, enabling the system to reason about free space, occluded regions, and geometry without relying on predefined object categories. This category-agnostic nature makes occupancy prediction particularly relevant to corner-case handling: an unusual object that a detector would miss may still produce a non-empty occupancy voxel, alerting downstream modules to an obstacle even when classification fails.

An emerging direction is world-model-based simulation. Methods such as GAIA-1 [13] and DriveDreamer [14] learn generative models that synthesize realistic future frames conditioned on ego-vehicle actions and scene context, offering a principled mechanism for corner-case data augmentation. However, fidelity and semantic diversity of generated scenes remain limited, and ensuring physically plausible sensor readings for rare configurations is an active challenge.

Research gap: Despite strong benchmark performance, BEV and occupancy methods have not been systematically evaluated on corner-case datasets such as CODA [15], and world models have not yet been used to generate semantically diverse corner cases at scale.

C. Scene Graph Generation

Scene graphs represent traffic environments as structured graphs where nodes correspond to objects (vehicles, pedestrians, infrastructure) and edges encode spatial, semantic, or functional relationships between them. This higher-level abstraction enables explicit reasoning about interactions, context, and causality, making scene graphs particularly suitable for relational tasks such as predicting traffic participant behavior or understanding right-of-way at intersections.

Jiang et al. [16] propose enhancing scene graph generation (SGG) by incorporating hierarchical relationships and commonsense knowledge, demonstrating improved relational reasoning through a structured integration of prior knowledge into the generation pipeline. Hu et al. [17] address the challenge of 3D SGG by introducing cross-modal alignment and adversarial learning, which improves the quality of scene graphs constructed from multi-modal sensory data. These works share a common insight: standard SGG methods suffer from biased relationship prediction, and external knowledge can mitigate this limitation.

In the context of autonomous driving specifically, Greve et al. [18] present Collaborative URBan Scene Graphs (CURB-SG), which leverage panoptic LiDAR data from multiple agents to build large-scale, hierarchical scene graphs of urban environments. Their method enables higher-order reasoning and efficient querying for various AD functions, demonstrating the practical value of scene graph representations for multi-agent driving systems. Zhang et al. [19] propose GraphAD, which introduces the Interaction Scene Graph (ISG) for end-to-end autonomous driving. By encoding prior knowledge about interactions into a directed graph model, GraphAD achieves state-of-the-art performance on the nuScenes benchmark across perception, prediction, and planning tasks while being more computationally efficient than attention-based alternatives. This result suggests that explicit graph-based interaction modeling is a viable and effective paradigm for AD.

Werby et al. [20] extend scene graphs into the open-vocabulary domain by constructing hierarchical 3D scene graphs enriched with natural language descriptions. Their approach enables robots to navigate based on language queries, which implicitly supports the detection of previously unseen objects through language grounding rather than fixed category labels. This line of work is particularly relevant to corner-case scenarios, where the objects encountered may not belong to any predefined category.

Research gap: Existing scene graph methods for autonomous driving rely on closed-set detectors and have not been extended to open-vocabulary settings. Furthermore, the robustness of graph representations under corner-case conditions — where critical nodes may be missing or misclassified — has not been systematically evaluated.

D. Open-Vocabulary Perception

Traditional perception models are trained on closed sets of predefined categories, which fundamentally limits their ability to handle corner cases involving novel objects. Open-vocabulary detection and segmentation methods address this limitation by leveraging vision-language models to recognize objects described by arbitrary text prompts, rather than being restricted to a fixed label set.

Zhu and Chen [21] provide an extensive survey of open-vocabulary detection and segmentation, tracing the evolution from early zero-shot methods to modern approaches based on contrastive vision-language pre-training. They identify the alignment between visual and textual feature spaces as the central technical challenge. Foundational models such as CLIP [22] have been instrumental in enabling this alignment by learning joint representations from large-scale image-text pairs. Building on these foundations, Grounding DINO [23] combines a DINO-based detector with grounded pre-training to achieve strong open-set object detection, allowing detection of arbitrary objects specified through text prompts.

These open-vocabulary approaches offer a fundamentally different strategy for corner-case handling: instead of enumerating all possible categories during training, the system can be prompted at inference time to detect novel objects based on their textual descriptions — directly relevant to autonomous driving, where the set of possible road objects is effectively unbounded.

However, as Zhu and Chen [21] note, the performance gap between open-vocabulary and fully supervised closed-set detectors remains substantial, particularly for fine-grained detection of small objects at long range, which is critical for AD applications. Furthermore, the computational cost of vision-language models poses challenges for real-time deployment: current state-of-the-art models require significant inference time compared to specialized closed-set detectors, making their integration into latency-sensitive AD pipelines non-trivial. Recent works on knowledge distillation and model

compression show promise in bridging this efficiency gap, but a practical real-time open-vocabulary detector for AD remains an active area of research.

Research gap: No existing work has integrated open-vocabulary perception directly into a scene graph pipeline for autonomous driving. The combination would allow the scene graph to represent novel, previously unseen objects — but the architectural, efficiency, and relationship-definition challenges of such integration remain unaddressed.

E. Corner-Case Detection and Generation

The formal characterization and systematic handling of corner cases has received increasing attention in recent years. Heidecker et al. [4] propose an application-driven conceptualization that organizes corner cases into multiple levels, from sensor-level anomalies to scenario-level deviations, providing a structured framework for analyzing perception failures. Bogdoll et al. [3] extend this perspective by examining corner cases from the machine learning model's viewpoint, identifying properties that are beneficial for their description, reproduction, and synthetic generation.

On the dataset side, Li et al. [15] introduce CODA, a real-world corner-case dataset constructed from major AD benchmarks (KITTI, nuScenes, ONCE). Their evaluation reveals a dramatic performance drop: state-of-the-art detectors achieve no more than 12.8% mean Average Recall on CODA, compared to significantly higher figures on standard benchmarks. Even open-world detectors fail to reliably identify the novel objects in CODA, underscoring the severity of the corner-case challenge.

Simulation-based approaches offer a complementary strategy for addressing corner-case data scarcity. Daoud et al. [24] present CornerSim, a virtualization framework that generates corner-case scenarios using the CARLA simulator by programmatically controlling environmental conditions (rain, fog, night), object placement (debris, animals, unusual vehicles), and traffic behaviors (aggressive lane changes, pedestrian jaywalking), enabling systematic evaluation under conditions that are difficult or dangerous to reproduce in the real world.

The broader data-centric perspective surveyed by Li et al. [5] provides context for how industry addresses the long-tail challenge. They describe closed-loop pipelines — Tesla's Fleet Learning and NVIDIA's MagLev — that automatically detect corner cases during real-world operation and incorporate them into retraining cycles, but note that such systems require massive fleets and infrastructure, limiting accessibility for smaller research groups.

The emerging use of multimodal large language models (MLLMs) for corner-case understanding represents a recent paradigm shift. Wang et al. [25] apply MLLMs to CODA, constructing the CODA-REC dataset with text descriptions and demonstrating that zero-shot LLM capabilities can enhance semantic understanding of unusual objects that conventional detectors miss. Similarly, NVIDIA's Alpamayo-R1 [26] incorporates a reasoning module that generates textual rationales before making driving decisions, targeting improved generalization in long-tail situations. These reasoning-augmented approaches suggest that integrating high-level semantic reasoning with low-level perception can bridge the gap between common and rare scenarios, though inference latency and reasoning reliability for safety-critical decisions remain open questions.

III. Problem Identification and Future Directions

A. Summary of Open Problems

The literature reveals a clear hierarchy of maturity. Standard-scenario perception — object detection, semantic segmentation, and end-to-end planning on curated benchmarks — has reached high performance. Structured scene representations, exemplified by GraphAD [19] and CURB-SG [18], demonstrate practical value for interaction reasoning. BEV feature extraction and 3D occupancy prediction provide geometrically rich scene descriptions that are becoming the dominant representation backbone.

By contrast, four critical problems remain substantially unsolved:

(P1) Open-world object detection under real-time constraints. The CODA benchmark [15] shows that state-of-the-art detectors achieve at most 12.8% mean Average Recall on novel objects. Open-vocabulary methods [21], [23] improve recognition breadth but operate at a fraction of the throughput required for embedded AD hardware. No existing method achieves both open-set generalization and real-time performance on corner-case-rich datasets simultaneously.

(P2) Scene graph construction for open-world scenes. All driving-focused scene graph methods [19], [18], [16] inherit the closed-set assumption of their underlying detectors. When an unusual object is not detected, it is entirely absent from the graph, and all downstream reasoning is silently compromised. The cascading failure mode introduced by this gap has not been characterized or mitigated in the literature.

(P3) Scalable and semantically diverse corner-case generation. CornerSim [24] and fleet-based pipelines [5] address scenario scarcity through simulation and data mining respectively, but neither produces sufficiently diverse semantic categories or provides ground-truth labels suitable for open-world evaluation. World-model approaches [13], [14] are promising but not yet validated against corner-case benchmarks.

(P4) Evaluation methodology for long-tail perception. Current AD benchmarks measure average performance, masking catastrophic failures on rare events. Heidecker et al.'s multi-level taxonomy [4] provides a conceptual framework, but a standardized quantitative evaluation protocol covering sensor-level, object-level, and scenario-level corner cases does not exist.

B. Proposed Research Directions

The analysis above motivates a focused research agenda centered on three concrete contributions.

Contribution 1: Open-vocabulary scene graph generation for autonomous driving. We propose to design a pipeline that couples a lightweight open-vocabulary detector — obtained through knowledge distillation from a large vision-language model — with an online scene graph builder. The key research questions are: (i) how to define relationship predicates for previously unseen object categories without re-training, leveraging language embeddings as a proxy for relational semantics; (ii) how to bound the inference latency of the combined system within the 50 ms budget typical of AD perception stacks; and (iii) how graph construction quality degrades as a function of the detector's open-world recall, using CODA [15] as a diagnostic benchmark.

Contribution 2: Corner-case-targeted data generation using world models. We propose to condition a world-model generator on rare object categories and unusual scene configurations to produce training and evaluation datasets that complement real-world collection. Validation will combine CornerSim [24]-style simulation for ground-truth control with real-sensor data from CODA for domain fidelity assessment. A key deliverable is a protocol for measuring the fidelity and semantic diversity of generated corner cases.

Contribution 3: Reasoning-augmented scene graph interpretation. Drawing on the demonstrated capabilities of MLLMs for corner-case description [25] and reasoning-before-action frameworks [26], we propose to attach a lightweight reasoning module to the scene graph output. This module would generate natural-language hazard descriptions for novel nodes, which can be audited by operators and used to trigger conservative planning behaviors. The key open problem is whether LLM-generated rationales can be made reliable enough for safety-relevant decisions without introducing unacceptable latency.

These three contributions are intended to be pursued incrementally: Contribution 1 addresses the core perception gap; Contribution 2 provides the evaluation infrastructure; and Contribution 3 closes the loop between perception and planning for corner-case scenarios.

REFERENCES

[1] J. Zhao, W. Zhao, B. Deng, Z. Wang, F. Zhang, W. Zheng, W. Cao, J. Nan, Y. Lian, and A. F. Burke, "Autonomous driving system: A comprehensive survey," *Expert Systems with Applications*, vol. 242, p. 122836, 2024.

[2] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, "End-to-end autonomous driving: Challenges and frontiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 10 164–10 183, 2024.

[3] D. Bogdoll, S. Uhlemeyer, K. Kowol, and J. M. Zöllner, "Corner cases in machine learning processes," *AI Perspectives and Advances*, vol. 1, no. 1, pp. 1–22, 2024.

[4] F. Heidecker, J. Breitenstein, K. Rösch, J. Löhdefink, M. Bieshaar, C. Stiller, T. Fingscheidt, and B. Sick, "An application-driven conceptualization of corner cases for perception in highly automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, 2021, pp. 644–651.

[5] L. Li, W. Shao, W. Dong, Y. Tian, Q. Zhang, K. Yang, and W. Zhang, "Data-centric evolution in autonomous driving: A comprehensive survey of big data system, data mining, and closed-loop technologies," 2024.

[6] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, 2021.

[7] L. Contreras, S. Jain, N. P. Bhatt, S. Banerjee, and E. Hashemi, "A survey on 3D object detection in real time for autonomous driving," *Frontiers in Robotics and AI*, vol. 11, p. 1212070, 2024.

[8] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, L. Lu, X. Jia, Q. Liu, J. Dai, Y. Qiao, and H. Li, "Planning-oriented autonomous driving," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 17 853–17 862.

[9] Z. Liu, H. Tang, A. Amini *et al.*, "BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2023, pp. 2774–2781.

[10] Z. Li, W. Wang, H. Li *et al.*, "BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in *Proc. European Conf. Computer Vision (ECCV)*, 2022, pp. 1–18.

[11] Y. Huang, W. Zheng, Y. Zhang *et al.*, "Tri-perspective view for vision-based 3D semantic occupancy prediction," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 9223–9232.

[12] Y. Wei, L. Zhao, W. Zheng *et al.*, "SurroundOcc: Multi-camera 3D occupancy prediction for autonomous driving," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023, pp. 21 729–21 740.

[13] A. Hu, L. Russell, H. Yeo *et al.*, "GAIA-1: A generative world model for autonomous driving," *arXiv preprint arXiv:2309.17080*, 2023.

[14] X. Wang, Z. Zhu, G. Huang *et al.*, "DriveDreamer: Towards real-world-driven world models for autonomous driving," in *Proc. European Conf. Computer Vision (ECCV)*, 2024, pp. 1–18.

[15] K. Li, K. Chen, H. Wang, L. Hong, C. Ye, J. Han, Y. Chen, W. Zhang, C. Xu, D.-Y. Yeung, X. Liang, Z. Li, and H. Xu, "CODA: A real-world road corner case dataset for object detection in autonomous driving," in *Proc. European Conf. Computer Vision (ECCV)*, 2022, pp. 406–423.

[16] B. Jiang, Z. Zhuang, S. S. Shivakumar, and C. J. Taylor, "Enhancing scene graph generation with hierarchical relationships and commonsense knowledge," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2025, pp. 8883–8894.

[17] Y. Hu, X. Zhou, C. Wang, W. Meng, and G. He, "3D scene graph generation with cross-modal alignment and adversarial learning," in *Proc. Int. Conf. Multimedia Retrieval (ICMR)*, 2025, pp. 487–496.

[18] E. Greve, M. Büchner, N. Vödisch, W. Burgard, and A. Valada, "Collaborative dynamic 3D scene graphs for automated driving," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2024, pp. 11 118–11 124.

[19] Y. Zhang, D. Qian, D. Li, Y. Pan, Y. Chen, Z. Liang, Z. Zhang, S. Zhang, H. Li, M. Fu, Y. Ye, Z. Liang, Y. Shan, and D. Du, "GraphAD: Interaction scene graph for end-to-end autonomous driving," in *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2025, pp. 2422–2430.

[20] A. Werby, C. Huang, M. Büchner, A. Valada, and W. Burgard, "Hierarchical open-vocabulary 3D scene graphs for language-grounded robot navigation," in *Proc. Robotics: Science and Systems (RSS)*, Delft, Netherlands, 2024.

[21] C. Zhu and L. Chen, "A survey on open-vocabulary detection and segmentation: Past, present, and future," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, pp. 8954–8975, 2024.

[22] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Machine Learning (ICML)*, 2021, pp. 8748–8763.

[23] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, and L. Zhang, "Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection," in *Proc. European Conf. Computer Vision (ECCV)*, 2024, pp. 38–55.

[24] A. Daoud, C. Bunel, and M. Guériau, "CornerSim: A virtualization framework to generate realistic corner-case scenarios for autonomous driving perception testing," in *Proc. 15th Int. Conf. Ambient Systems, Networks and Technologies (ANT)*, vol. 238, 2024, pp. 184–191.

[25] Y. Wang, R. Zheng, Y. Peng, J. Zhao, C. Wang, and R. Du, "Empowering corner case detection in autonomous vehicles with multimodal large language models," *IEEE Internet of Things Journal*, 2024.

[26] NVIDIA, "Alpamayo-R1: Bridging reasoning and action prediction for generalizable autonomous driving in the long tail," *CoRR*, vol. abs/2511.00088, 2025.

Global Optimization and Regularity in Discrete Dynamic Systems over Extremal Algebras

[1]*Nikola TOMAŠČÍKOVÁ (2nd year),*
Supervisor: [2]*Helena MYŠKOVÁ*

[1,2]Dept. of Mathematics and Theoretical Informatics, FEI, Technical University of Košice, Slovak Republic

[1]nikola.tomascikova@tuke.sk, [2]helena.myskova@tuke.sk

***Abstract*—This paper focuses on discrete dynamic systems modeled using extremal algebras with an emphasis on global optimization problems in max-plus algebra and solvability of linear systems in max-min algebra under uncertainty. The current state of research in these areas is reviewed, highlighting existing results on global optimization with linear and non-negativity constraints, interval extensions of max-plus systems, and strong regularity of max-min matrices. Particular attention is paid to algorithmic aspects and structural conditions that preserve unique solvability in the presence of imprecise data.**

***Keywords*—global optimization, max-min algebra, max-plus algebra, regularity**

I. Introduction

Discrete dynamic systems represent an important class of mathematical models used to describe processes whose evolution is governed by synchronization, temporal dependencies, and interactions between events. Such systems frequently arise in production planning, transportation networks, communication infrastructures, and computer-controlled processes, where timing constraints play a dominant role. Classical modeling approaches based on conventional linear algebra often fail to adequately capture these characteristics, motivating the use of alternative algebraic frameworks.

Extremal algebras, particularly max-plus and max-min algebra, provide a natural mathematical formalism for modeling discrete event systems. By replacing standard arithmetic operations with maximum, minimum, and addition, these algebras allow synchronization mechanisms and temporal delays to be expressed in a compact and analytically tractable way. Linear systems formulated in extremal algebras enable the analysis of steady states, long-term behavior, and performance bounds of discrete dynamic systems.

In practical applications, system parameters are rarely known exactly. Measurement errors, incomplete information, and environmental variability lead to uncertainty that must be explicitly taken into account. This motivates the study of systems with inexact and interval-valued data, where classical solvability and optimization results no longer apply directly. In this context, global optimization in max-plus algebra and solvability of linear systems in max-min algebra play a crucial role, particularly when extended to interval settings.

II. Literature Review and Analysis

Extremal algebra provides a mathematical framework for modeling systems whose behavior is governed by synchronization, delays, and bottleneck effects. Its foundations were established in early works on max-plus and max-min algebras, where classical arithmetic operations are replaced by extremal operators [1], [2]. A systematic treatment of discrete event systems using max-plus algebra was presented in the monograph by Baccelli et al. [3], which demonstrated that synchronization and timing phenomena can be described by linear equations over idempotent semirings. Subsequent developments clarified the algebraic, graph-theoretic, and algorithmic properties of such systems [4], [5], [6].

Max-plus algebra has since become a standard tool for the analysis of discrete dynamic systems. Comprehensive treatments of max-linear systems and their computational aspects are provided in [7], [8]. These works show that long-term system behavior is closely related to eigenvalues and eigenvectors of max-plus matrices, which correspond to cycle times and steady-state regimes. Extensions to systems with inexact data highlight that even small uncertainties in parameters can significantly affect system dynamics [9].

Global optimization in the context of extremal algebra represents a distinct research direction that differs substantially from classical optimization theory. General deterministic approaches to global optimization are surveyed in [10], [11], [12], [13], [14]. Within extremal algebra, optimization problems typically involve minimizing max-plus linear mappings over convex sets defined in the classical sense. Tao et al. [15], [16] investigated solvability conditions and structural properties of such problems, demonstrating that the existence of a global optimum depends on the interplay between algebraic structure and feasible set geometry.

A key challenge arises when non-negativity constraints are imposed, reflecting the interpretation of decision variables as time quantities. These constraints lead to new solvability phenomena that are absent in unconstrained problems [16]. Recent work has shown that, under suitable structural assumptions, global optimization problems in max-plus algebra admit unique optimal solutions that can be computed algorithmically [17]. These results form an important basis for further extensions to uncertain and interval-valued systems.

The incorporation of interval data is essential for modeling uncertainty in real-world systems. Interval max-plus systems of linear equations were studied in [18], [19], where solvability conditions were derived for systems with bounded coefficients. Further advances addressed structural properties of interval matrices and their regularity [20], [21]. Interval extensions of global optimization problems in max-plus algebra were recently formulated, providing necessary and sufficient con-

ditions for the existence of a feasible realization admitting a global optimum [17].

Parallel to max-plus algebra, max-min algebra has been developed as a natural framework for systems with upper bounds and capacity constraints. Early studies established connections between unique solvability and matrix properties [22]. The notion of strong regularity emerged as a central concept guaranteeing unique solvability of max–min linear systems [23]. Algorithmic criteria based on matrix permutations and trapezoidal structures enable efficient verification of this property.

Recent research has extended strong regularity to interval max-min matrices, introducing concepts such as possible and universal strong regularity [24], [25]. These results provide decision criteria for solvability in the presence of uncertainty and form a foundation for polynomial-time algorithms. However, despite substantial theoretical progress, the computational complexity of decision procedures for large-scale interval systems remains a limiting factor.

From a broader perspective, the literature reveals a clear gap between theoretical characterization and practical algorithmic efficiency. While solvability and optimization conditions in extremal algebras are now well understood, their extension to interval data introduces combinatorial complexity that necessitates further research. This motivates the present work, which aims to refine algorithmic approaches to global optimization and strong regularity in extremal algebras, with particular emphasis on tractable subclasses and interval uncertainty.

III. Background and Progress

The initial phase of the research focused on establishing a solid theoretical foundation in extremal algebraic structures and discrete dynamic systems. Fundamental properties of max-plus and max-min algebra were studied, including algebraic operations, matrix representations, and their interpretation in discrete event modeling. Particular attention was paid to linear systems and to the role of principal solutions, which form the basis for solvability analysis in both algebraic settings.

Subsequently, global optimization problems in max-plus algebra were analyzed in detail. The study first addressed the basic global optimization problem with linear constraints, followed by extensions incorporating non-negativity constraints motivated by the temporal interpretation of decision variables. For each problem variant, necessary and sufficient solvability conditions were examined, and algorithmic procedures were formulated to decide the existence of a global optimum and to compute it explicitly when it exists. These results demonstrate that, despite the nonlinearity of max-plus mappings in the classical sense, global optimization problems retain a structured and analyzable form.

The research then progressed toward systems with interval-valued data, which naturally arise when system parameters are subject to uncertainty. Interval matrices were introduced as a formal model of imprecise coefficients, and their impact on solvability and optimization was investigated. In this context, criteria for row dependence and regularity of interval matrices were studied, enabling the extension of solvability results from exact to interval settings. Algorithmic decision procedures were developed to determine whether a feasible realization of an interval system admits a global optimal solution.

In parallel, linear systems in max-min algebra were examined with emphasis on strong regularity of system matrices. The relationship between strong regularity and unique solvability was analyzed, and algorithmic criteria based on structural matrix properties were studied as effective tools for verifying this condition. These results were subsequently extended to interval max-min matrices, introducing refined notions of regularity that account for imprecise data while preserving decidability.

Overall, the completed work established a coherent theoretical and algorithmic framework integrating discrete dynamic systems, extremal algebras, global optimization, and interval uncertainty. This framework serves as a foundation for further research aimed at extending optimization and solvability concepts to more general extremal algebraic structures while maintaining computational feasibility.

IV. Prospective Extensions in Extremal Algebra

The results obtained so far confirm that extremal algebras constitute a powerful and flexible framework for the analysis of discrete dynamic systems with synchronization, constraints, and uncertainty. A key advantage of this framework lies in the existence of algebraic notions of linearity, principal solutions, and structural matrix properties, which together enable precise solvability and optimization analysis even in non-classical settings. These features provide a natural starting point for further theoretical extensions.

A primary research direction concerns the global optimization problem formulated in max-plus algebra. Existing results show that such problems admit well-defined solvability conditions and constructive solution methods when the system parameters are known exactly. Importantly, recent developments indicate that these properties can be preserved under interval uncertainty, provided that suitable structural assumptions are satisfied. This suggests that the global optimization framework in max-plus algebra is inherently robust with respect to imprecise data.

Beyond interval extensions, the algebraic nature of the optimization problem points toward its applicability in other extremal algebras. Many extremal algebraic structures share core properties such as idempotency, monotonicity, and order-based operations, which are essential for defining linear mappings and optimization criteria. This creates a realistic potential for extending global optimization concepts beyond max-plus algebra, while retaining meaningful solvability notions and algorithmic decision procedures. Such extensions would contribute to a more unified theory of optimization in extremal algebraic systems.

A second major research direction focuses on solvability of linear systems in max-min algebra. Strong regularity has emerged as a fundamental concept ensuring unique solvability and stability of solutions. Its algorithmic character, based on structural properties of system matrices, makes it particularly attractive for applications. However, classical strong regularity is formulated for exact data and does not fully capture uncertainty present in real-world systems.

The extension of strong regularity to imprecise and interval data provides an opportunity to refine solvability concepts without sacrificing computational feasibility. In this context, strong b-regularity represents a meaningful generalization that incorporates bounded uncertainty directly into the definition of regularity. The structure of max-min algebra suggests that such extensions can preserve polynomial-time testability,

which is essential for both theoretical relevance and practical applicability.

Taken together, these research directions highlight a common underlying theme: the possibility of systematically extending optimization and solvability concepts in extremal algebras while maintaining a balance between theoretical rigor and algorithmic efficiency. The potential of this approach lies not only in addressing uncertainty, but also in strengthening the connection between algebraic structure and computational tractability. This creates a coherent framework for future research on discrete dynamic systems modeled using extremal algebras.

V. Conclusion and Future Work

This work has focused on the analysis of discrete dynamic systems modeled using extremal algebras, with particular emphasis on max-plus and max-min algebra in the presence of imprecise and interval data. The theoretical and algorithmic results discussed so far demonstrate that key solvability and optimization properties can be preserved even under uncertainty, provided that appropriate structural conditions are satisfied.

The future research will build directly on these foundations. The primary objective is to extend global optimization problems formulated in max-plus algebra to interval data and to other types of extremal algebras, while maintaining clear solvability criteria and constructive solution methods. A further goal is to develop and analyze extended notions of regularity in max-min algebra, in particular strong $\boldsymbol{b}$-regularity with imprecise data, and to design decision procedures that are testable in polynomial time. By pursuing these directions, the research aims to deepen the theoretical understanding of extremal algebraic systems and to strengthen their applicability to real-world problems involving uncertainty, synchronization, and constrained dynamics.

References

[1] K. Zimmermann, *Extremální algebra*. Útvar vědeckých informací Ekonomického ústavu ČSAV, 1976.

[2] R. A. Cuninghame-Green, "Minimax algebra and applications," *Fuzzy Sets and Systems*, vol. 41, no. 3, pp. 251–267, 1991.

[3] F. Baccelli, G. Cohen, G. J. Olsder, and J. Quadrat, *Synchronization and Linearity - An Algebra for Discrete Event Systems*. Wiley New York, 1992, vol. 1.

[4] S. Gaubert and M. Plus, "Methods and applications of (max,+) linear algebra," in *Annual symposium on theoretical aspects of computer science*. Springer, 1997, pp. 261–282.

[5] M. Gondran and M. Minoux, *Graphs, Dioids and Semirings*. Springer, 2008.

[6] J. S. Golan, *Semirings and Their Applications*. Springer, 1999.

[7] P. Butkovič, *Max-linear systems: theory and algorithms*. Springer Science & Business Media, 2010.

[8] B. Heidergott, G. J. Olsder, and J. Van Der Woude, *Max Plus at work: modeling and analysis of synchronized systems: a course on Max-Plus algebra and its applications*. Princeton University Press, 2014, vol. 48.

[9] H. Myšková and J. Plavka, "Max-plus steady states in discrete event dynamic systems with inexact data," *Discrete Event Dynamic Systems*, vol. 32, no. 3, pp. 521–538, 2022.

[10] I. E. Grossmann and L. T. Biegler, "Part ii. future perspective on optimization," *Computers & Chemical Engineering*, vol. 28, no. 8, pp. 1193–1218, 2004.

[11] A. Neumaier, "Complete search in continuous global optimization and constraint satisfaction," *Acta numerica*, vol. 13, pp. 271–369, 2004.

[12] C. A. Floudas and C. E. Gounaris, "A review of recent advances in global optimization," *Journal of Global Optimization*, vol. 45, no. 1, pp. 3–38, 2009.

[13] B. Hartke, "Global optimization," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 6, pp. 879–887, 2011.

[14] R. Horst and H. Tuy, *Global optimization: Deterministic approaches*. Springer Science & Business Media, 2013.

[15] Y. Tao, G. Liu, and W. Chen, "Globally optimal solutions of max–min systems," *Journal of Global Optimization*, vol. 39, no. 3, pp. 347–363, 2007.

[16] Y. Tao and C. Wang, "Global optimization for max-plus linear systems and applications in distributed systems," *Automatica*, vol. 119, p. 109104, 2020.

[17] H. Myšková and J. Plavka, "Interval global optimization problem in max-plus algebra," *Linear Algebra and its Applications*, vol. 714, pp. 45–63, 2025.

[18] H. Myšková, "Interval systems of max-separable linear equations," *Linear Algebra and its Applications*, vol. 403, pp. 263–272, 2005.

[19] ——, "Interval max-plus systems of linear equations," *Linear algebra and its applications*, vol. 437, no. 8, pp. 1992–2000, 2012.

[20] H. Myšková and J. Plavka, "Regularity of interval max-plus matrices," *Linear Algebra and its Applications*, vol. 680, pp. 28–44, 2024.

[21] J. Plavka and S. Sergeev, "X-simple image eigencones of tropical matrices," *Linear Algebra and Its Applications*, vol. 507, pp. 169–190, 2016.

[22] K. Cechlárová, "Unique solvability of max-min fuzzy equations and strong regularity of matrices over fuzzy algebra," *Fuzzy sets and systems*, vol. 75, no. 2, pp. 165–177, 1995.

[23] M. Gavalec and J. Plavka, "Strong regularity of matrices in general max-min algebra," *Linear Algebra and its Applications*, 2003.

[24] H. Myšková and J. Plavka, "Regularity of interval fuzzy matrices," *Fuzzy Sets and Systems*, vol. 463, p. 108478, 2023.

[25] M. Gavalec, J. Ramík, and K. Zimmermann, "Interval eigenproblem in max-min algebra," in *Decision Making and Optimization: Special Matrices and Their Applications in Economics and Management*. Springer, 2014, pp. 163–181.

Research on the possibilities of increasing the integration of renewable energy sources into the electric power system

[1]*Kristián ELIÁŠ (2nd year)*
Supervisor: [2]Ľubomír BEŇA

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]kristian.elias@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract*—As a result of the energy policy of the European Union, there is currently a rapid expansion of the connection of photovoltaic systems to electric networks across Europe. However, the operation of photovoltaic systems is characterized by impact on the distribution network. This paper therefore focuses on determining the influence of photovoltaic systems on the operation of the distribution network. For modeling this influence, a model low-voltage distribution network is proposed and implemented in the EA-PSM simulation software. Four model scenarios are proposed to determine the impact of photovoltaic systems on the network, and the EN 50160 standard is used in the evaluation of this impact. The modeling results thus help to understand the challenges associated with the massive integration of photovoltaic systems into low-voltage distribution networks.**

***Keywords*—distribution system, power quality, photovoltaic systems, renewable energy sources.**

I. Introduction

Following the Paris Agreement of 2015, which legally binds its signatories to combat climate change [1], the European Commission presented the European Green Deal (EGD) in 2019, whose main aim is to achieve climate neutrality by 2050 [2]. Part of the EGD is also a mid-term target for 2030, which is to reduce greenhouse gas emissions by at least 55% compared to 1990 levels. These targets became legally binding for all Member States of the European Union (EU) through the adoption of the European Climate Law [3]. To support the achievement of the 2030 target, the European Commission introduced in 2023 a legislative package known as Fit for 55. This package includes measures focused on the development of renewable energy sources (RES), improving energy efficiency, increasing the energy performance of buildings, and the development of electromobility [4]. To accelerate the transition to clean energy, the European Commission also presented the REPowerEU plan in 2022, the key element of which is the rapid and large-scale deployment of RES, particularly solar energy. Within the REPowerEU plan, the objective is therefore to have at least 600 GW of solar photovoltaic connected to the electrical grids in the EU by 2030 (according to the most recent data, 406 GW were connected to the EU grids in 2025 [5]). In particular, the vast, so far unused potential of solar energy on building rooftops across the EU should begin to be utilized. According to some estimates, rooftop photovoltaics could cover up to 25% of electricity consumption in the EU. For this reason, an obligation to install solar installations on buildings in the EU will be gradually introduced, where it is technically suitable and economically and functionally feasible [6]. A key factor for the rapid and large-scale integration of photovoltaic (PV) systems into the EU electrical grids is, above all, an appropriate energy infrastructure. It is estimated that approximately 40% of distribution systems in the EU are more than 40 years old, and therefore their modernization will be necessary. The estimated investment costs required for the modernization of distribution networks in the EU by 2030 alone range from 375 to 425 billion euros [7].

For planning the development of distribution networks, it is therefore necessary to choose an appropriate strategy that ensures the networks are prepared to absorb as many PV systems as possible without compromising operational reliability and safety. It is thus essential to understand the impact of PV system operation on the distribution network, specifically on its quality parameters (voltage magnitude and voltage unbalance) and operational parameters (currents and system losses). In a previous paper presented at the SCYR conference [8], based on the analysis of data from smart meters installed at consumption points with connected single-phase and three-phase PV systems, I demonstrated their impact on voltage magnitude at the point of connection to the distribution system. However, due to the absence of additional data, it was not possible to determine the comprehensive impact of PV systems also on other parameters of the distribution network. In this year I therefore focused on modeling a low-voltage distribution network with installed PV systems and analyzing their impact on selected parameters of the distribution network.

II. Methodology

For the analysis of the impact of single-phase and three-phase PV systems, a classical deterministic approach was used, in which attention is focused on the worst-case operating scenarios of the distribution network (maximum possible PV output, connection of single-phase PV systems to only one phase, etc.). The model demonstration network replicates the structure of real distribution networks in

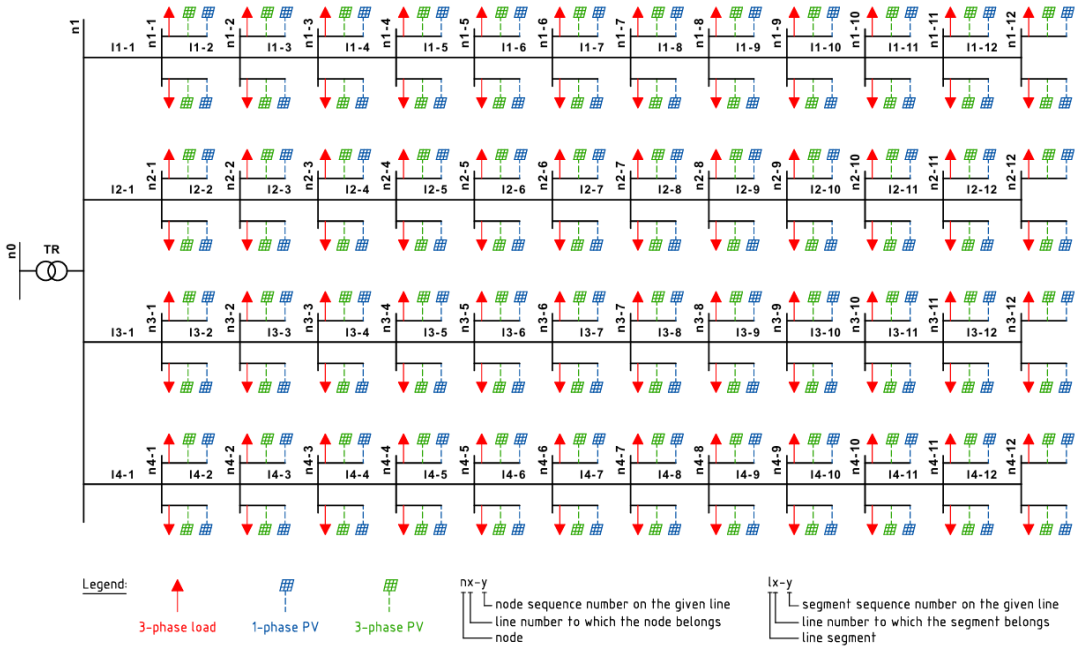


Fig. 1. General topology of the modeled network

Slovakia and is composed of elements used in the construction of distribution networks. The general schematic of the model low-voltage distribution network is shown in Fig. 1. The model distribution network consists of four lines originating from node n1. Each line has a length of 600 m and contains 12 nodes evenly distributed along the line. At each node, there are two consumption points where a three-phase symmetrical load of 3.464 kVA with a power factor of 1 is connected. The voltage at the slack node n0 is 22.66 kV. The individual lines are constructed using cables of different types and cross-sectional areas. The parameters of the lines are listed in Table I.

TABLE I
PARAMETERS OF LINES

Line	Type	R [Ω/km]	X [Ω/km]	I_n [A]
l1	NFA2X 4x95	0,32	0,083	240
l2	NFA2X 4x120	0,25	0,082	280
l3	NAYY-J 4x150	0,206	0,080	254
l4	NAYY-J 4x240	0,129	0,080	332

In addition to the loads, single-phase or three-phase PV systems were also connected to the consumption points in the individual model scenarios. A total of four model scenarios were simulated, one of which serves as the reference scenario. An overview of the simulated scenarios is provided in Table II.

TABLE II
OVERVIEW OF MODELED SCENARIOS

Scenario	Description
A0	in the consumption points, only loads are connected (i.e., no PV systems are connected)
A1	in each consumption point, a single-phase PV system (P = 3,68 kW, cos φ = − 0,97) connected to phase A is installed
A2	in each consumption point, a three-phase PV system (P = 11 kW, cos φ = − 0,97) is connected
A3	at each node of the distribution system, in one consumption point a single-phase PV system (P = 3,68 kW, cos φ = − 0,97) connected to phase A is installed, and in the second consumption point a three-phase PV system (P = 11 kW, cos φ = − 0,97) is connected

Single-phase and three-phase PV systems are operated with a constant power factor cos φ = − 0.97 in the individual modeled scenarios, where they draw reactive power from the grid. The operating mode of the PV systems and the specific power factor value were selected based on the rules [9] for operating sources in low-voltage distribution networks established by the distribution company VSD, a. s.

The commercial software EA-PSM was used to calculate the steady-state operation, as it enables a comprehensive three-phase analysis of unbalanced electrical networks. When analyzing the impact of PV systems on the distribution network, the following parameters were monitored:

- voltage magnitude, which according to the EN 50160:2023 standard [10] should be within the range of 230 V ± 10%;
- voltage unbalance, which according to the EN 50160:2023 standard [10] should be within the range of 0% to 2%; and
- currents flowing through the lines, for which the limit value is determined by the maximum permissible current-carrying capacity of the cable (see Table I).

III. MODELING RESULTS

This chapter contains a graphical interpretation of the results of modeling the impact of PV systems on the distribution network. The effect of PV system operation on voltage magnitude is shown in Fig. 2. From Fig. 2 it can be seen that in none of the modeled scenarios with installed PV systems are the voltage magnitudes within the required limits at all nodes on any line.

In scenario A1, the voltage in phase B exceeded the maximum permissible value of 253 V at all nodes on all lines. The maximum voltage value in phase B in scenario A1 was 298 V on line 1 (20% above the limit), 293 V on line 2 (17% above the limit), 290 V on line 3 (16% above the limit), and 286 V on line 4 (14% above the limit). However, the phase voltages in phases A and C were within the permissible limits

at all nodes on all lines in scenario A1.

In scenario A2, the phase voltages at individual nodes on the lines had identical values, as only three-phase PV systems were connected to the originally symmetric network in this scenario. The highest number of nodes where the maximum permissible voltage value of 253 V was exceeded occurred on lines 1 and 2, where the voltage exceeded the maximum permissible value in 10 out of 12 nodes. On line 3, the maximum permissible voltage value was exceeded in 9 nodes, and on line 4 in 5 nodes. The highest voltage value was 268 V on line 1 (7% above the limit), 263 V on line 2 (4% above the limit), 260 V on line 3 (3% above the limit), and 254 V on line 4 (0,4% above the limit).

In scenario A3, the phase voltage magnitudes at individual nodes of the distribution network differed, similarly to scenario A1, as a result of the presence of single-phase PV systems. In scenario A3, the voltages in phase B increased to such an extent that the voltage in phase B was outside the permissible limits at every node on every line. The highest voltage value in phase B in scenario A3 was 280 V on line 1 (12% above the limit), 276 V on line 2 (10% above the limit), 273 V on line 3 (9% above the limit), and 268 V on line 4 (7% above the limit). In scenario A3, the voltage in phase A also exceeded the maximum permissible value, specifically in 5 out of 12 nodes on line 1. The highest voltage value in phase A on this line was 254 V, which is 0,4% above the limit. However, on lines 2 to 4, the voltage in phase A was within the permissible limits at all nodes. The voltage in phase C was also within the permissible limits in scenario A3, remaining within the allowed bounds at all nodes on all lines.

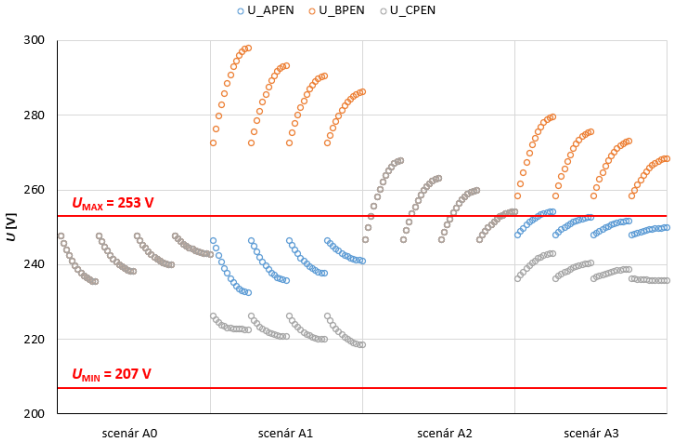


Fig. 2. Summary graphical representation of phase voltages at all network nodes for the individual modeled scenarios

The voltage unbalance in the individual modeled scenarios is illustrated in Fig. 3. From Fig. 3 it can be seen that voltage unbalance occurred only in scenarios A1 and A3, when single-phase PV systems were connected to the distribution network. In scenario A1, the voltage unbalance across the entire distribution network reached its highest value at a node on the line with the smallest conductor cross-section, i.e., line 1, where it reached 7,61% at the last node on the line (5,61% above the maximum permissible value). On line 2, the highest voltage unbalance value in scenario A1 was 6,78%, on line 3 it was 6,28%, and on line 4 it was 5,54%. The results show that in scenario A1, the voltage unbalance exceeded the permissible limits at every node in the entire distribution network due to the high penetration of single-phase PV systems connected to only one phase of the distribution network.

In scenario A3, where half as many single-phase PV systems were connected in the distribution network, the voltage unbalance at individual nodes of the distribution network is visibly lower. Nevertheless, the voltage unbalance still exceeds the maximum permissible value on lines 1 and 2 at all nodes, and on lines 3 and 4 at 11 out of 12 nodes. Similarly to scenario A1, the voltage unbalance across the network was highest at the last node on line 1, reaching 3,82%. On line 2, the highest voltage unbalance value was 3,41%, on line 3 it was 3,16%, and on line 4 it was 2,81%.

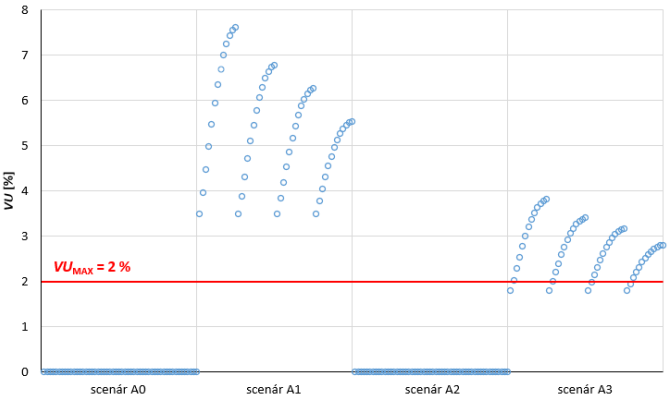


Fig. 3. Summary graphical representation of voltage unbalance at all network nodes for the individual modeled scenarios

Fig. 4 shows the current magnitudes flowing through the line sections in the individual modeled scenarios. Currents marked with a cross represent currents exceeding the maximum permissible current-carrying capacity of the line conductors.

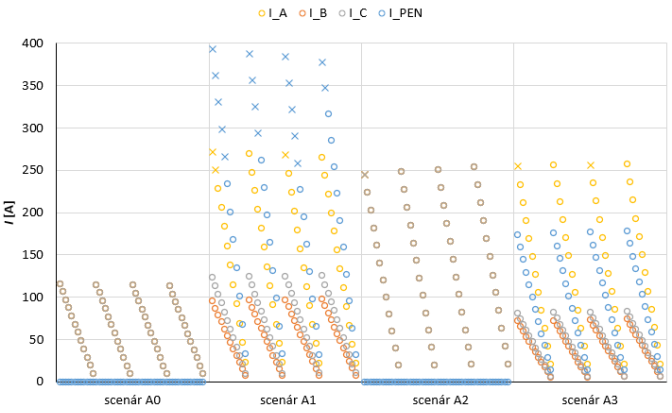


Fig. 4. Summary graphical representation of currents flowing through line sections for the individual modeled scenarios

From Fig. 4 it can be seen that the largest number of line sections was clearly overloaded in operating scenario A1. In this operating scenario, the highest currents flowed through the PEN conductors, as a result of significant asymmetry caused by the presence of single-phase PV systems. In modeled scenario A3, when half as many single-phase PV systems were connected to the distribution network compared to scenario A1, no overload of the PEN conductor occurred in any section of the distribution network. In scenario A3, only phase conductor A was overloaded, specifically in the first section of lines 1 and 3. In section l1-1, the current flowing through phase conductor A was 255,1 A, exceeding the conductor's rated current by 6,3%. In section l3-1, the current was 256,9 A, representing an exceedance of the nominal

current value of the line 3 conductors by approximately 1%.

In model scenario A2, only a single section was overloaded across the entire distribution network, namely section l1-1, through which a current of 245,7 A flowed (exceeding the maximum permissible value by 2,4%).

In scenario A1, overloading occurred in certain sections on all lines of the distribution network. On the line with the largest conductor cross-section, i.e., line 4, the PEN conductor was overloaded in sections l4-1 and l4-2. On line 3, the PEN conductor was overloaded in sections l3-1 to l3-5, and also phase conductor A in section l3-1. On line 2, the PEN conductor was overloaded in the first four sections of the line. On line 1, the PEN conductor was overloaded in the first five sections of the line, and phase conductor A in the first two sections. The maximum current value flowing through the conductors of line 1 was 393,7 A, exceeding the nominal current of the line conductors by 64%. The maximum current value flowing through the conductors of line 2 was 388,3 A, exceeding the nominal current of the line conductors by 39%. The maximum current value flowing through the conductors of line 3 was 384,8 A, exceeding the nominal current of the line conductors by 51%. The maximum current value flowing through the conductors of line 4 was 378,7 A, exceeding the nominal current of the line conductors by 14%.

IV. Conclusion and Future Plans

The modeling results show that a high penetration of PV systems in the distribution network can, under certain conditions, compromise the operational reliability and safety of the distribution network. Deteriorated power quality in the distribution network may compromise the proper functioning of electrical devices supplied by the network or even lead to their permanent damage. The results also indicate that PV systems can cause line overloading, which in practice would be detected by protection systems and result in interruption of the electricity supply to end consumers. Such events are undesirable in the operation of low-voltage distribution networks, and therefore the development and operation of distribution networks must be adapted so that, taking into account the expected increasing penetration of PV systems in distribution networks, the probability of undesirable events threatening the reliability and safety of the distribution network is minimized.

In the next phase, I will therefore focus on designing measures that minimize the negative impact of PV systems on the distribution network and ensure the safety of its operation under an increasing number of connected decentralized electricity generation units. These measures will then be implemented into the developed low-voltage distribution network model and their effectiveness will be verified.

Acknowledgment

This paper was supported by the Science Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under the contract VEGA 1/0532/25 and by the Slovak Research and Development Agency under contracts No. APVV-19-0576 and APVV-21-0312.

References

[1] "Paris Agreement on climate change." consilium.europa.eu. Accessed: Feb. 5, 2026. [Online]. Available: https://www.consilium.europa.eu/en/policies/paris-agreement-climate/

[2] "Európska zelená dohoda." minzp.sk. Accessed: Feb. 5, 2026. [Online]. Available: https://www.minzp.sk/klima/europska-zelena-dohoda/

[3] "European Climate Law." commission.europa.eu. Accessed: Feb. 5, 2026. [Online]. Available: https://climate.ec.europa.eu/eu-action/european-climate-law_en

[4] "Fit for 55." consilium.europa.eu. Accessed: Feb. 5, 2026. [Online]. Available: https://www.consilium.europa.eu/en/policies/fit-for-55/

[5] "Solar energy." consilium.europa.eu. Accessed: Feb. 5, 2026. [Online]. Available: https://energy.ec.europa.eu/topics/renewable-energy/solar-energy_en

[6] European Commission, "EU Solar Energy Strategy," Brussels, COM(2022) 221 final, May 18, 2022. Accessed: Feb. 5, 2026. [Online]. Available: https://eur-lex.europa.eu/resource.html?uri=cellar:516a902d-d7a0-11ec-a95f-01aa75ed71a1.0001.02/DOC_1&format=PDF

[7] European Commission, "Grids, the missing link - An EU Action Plan for Grids," Brussels, COM(2023) 757 final, Nov. 28, 2023. Accessed: Feb. 5, 2026. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52023DC0757

[8] K. Eliáš, "Research on the possibilities of increasing the integration of renewable energy sources into the electric power system," in *25^{th} Scientific Conf. of Young Researchers*, 2025, pp. 63-66.

[9] VSD, a. s., Košice, Slovakia. *TECHNICKÉ PODMIENKY PREVÁDZKOVATEĽA DISTRIBUČNEJ SÚSTAVY – VÝCHODOSLOVENSKÁ DISTRIBUČNÁ, A.S.,* (2024). Accessed: Mar. 30, 2026. [Online]. Available: https://www.vsds.sk/mdoc/dso.B6000.A/doc/20240901_VSD_Technicke_Podmienky_PDS.pdf

[10] *Voltage characteristics of electricity supplied by public electricity networks*, EN 50160, 2023.

Autonomous Knowledge Synthesis: Multi-Agent Methodology for Systematic Scientific Literature Analysis

[1]*Michal STROMKO (1st year),*
Supervisor: [2]*Peter BEDNÁR*

[1,2]Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovak Republic

[1]michal.stromko@tuke.sk, [2]peter.bednar@tuke.sk

***Abstract*—The exponential growth of scientific publications poses a significant challenge for researchers attempting to maintain a comprehensive understanding of the state-of-the-art. Traditional manual review methods and standard keyword-based retrieval are increasingly insufficient for handling the scale of modern research streams. While Large Language Models (LLMs) offer potential for automation, monolithic approaches often suffer from hallucinations and a lack of critical reasoning depth. This paper reviews the current landscape of automated literature analysis and proposes a novel methodological framework for "Autonomous Knowledge Synthesis." The proposed architecture shifts from single-model prompting to a Multi-Agent System (MAS) approach, employing a coalition of specialized agents, including roles for relevance assessment, methodology critique, and innovation detection. By decomposing the literature review process into discrete, interactive tasks, this methodology aims to enhance the objectivity, scalability, and reproducibility of scientific synthesis. The paper outlines the architectural design of the MAS and discusses its potential to transform how researchers discover and integrate new knowledge.**

***Keywords*—Multi-Agent Systems, Adversarial Agent Verification, Large Language Models, Natural Language Processing, Automated Literature Review, Hallucination Mitigation, Systematic Review Automation.**

I. Introduction

The volume of scientific production in the fields of Artificial Intelligence (AI) and Natural Language Processing (NLP) has been growing at an exponential rate over the last decade. This surge in information poses a critical barrier not only for the academic community but also increasingly for industry professionals, such as Data Scientists and Machine Learning engineers. For these practitioners, the ability to rapidly identify and extract the latest methodologies is crucial for successful technology transfer from academia to real-world business solutions.

Traditional search methods based on keywords and metadata filtering on platforms like *Web of Science* or *Scopus* are losing their effectiveness. In an environment where thousands of relevant papers are published monthly, researchers face "cognitive overload," leading to the risk of overlooking critical innovations.

Although Large Language Models (LLMs) have demonstrated the potential to democratize access to expert knowledge [1], their direct application via basic prompting techniques (e.g., "zero-shot prompting") to complex scientific tasks reveals significant limitations. Monolithic models often exhibit context loss when processing long technical documents, methodological shallowness, and a tendency toward hallucinations. A simple "ask the model" approach fails to simulate the rigorous process of critical scientific peer review required for high-quality knowledge synthesis.

Recent advances suggest that the solution lies in the transition toward sophisticated multi-agent ecosystems (MAS). The work of Park et al. [2] demonstrates that interactive simulations of human behavior—generative agents—can maintain consistent roles and memory. This concept is further enhanced by the ReAct approach (Reasoning + Acting) proposed by Yao et al. [3], which enables models to synergize reasoning with action execution.

This paper addresses the gap between the need for scalable literature review and the limitations of current LLM tools by proposing a methodology for **Autonomous Knowledge Synthesis**, which we define as the process of critically filtering, extracting, and structuring relevant findings from large scientific corpora, enabling researchers to focus exclusively on publications applicable to their research domain. To this end, we aim to answer two core research questions:

- **RQ1:** Can a multi-agent architecture with specialized and opposing analytical personas achieve a higher number of evaluated research dimensions per paper compared to standard single-model summarization approaches?
- **RQ2:** Does the implementation of cross-agent verification significantly reduce the occurrence of factual hallucinations in extracted technical parameters?

II. State of the Art

This section surveys the current state of knowledge in automated literature analysis, critically evaluating existing approaches and identifying open challenges that motivate the proposed methodology. The automation of scientific knowledge synthesis has evolved from simple bibliometric tools to advanced cognitive architectures. This section explores the transition from traditional retrieval methods to autonomous multi-agent systems.

A. Evolution of Systematic Literature Reviews

Traditional systematic literature reviews (SLR) are highly dependent on manual screening within databases such as Web

of Science and Scopus. While these platforms provide high-quality peer-reviewed metadata, the sheer volume of output in the AI era makes manual synthesis nearly impossible. Early automation attempts focused on keyword extraction and citation mapping, but as D'Aniello et al. [4] argue, these methods lack the semantic depth required to understand methodological nuances. The need for tools that can "read" and "critique" content, rather than just index it, has become a primary requirement for modern R&D departments.

Systematic literature reviews typically follow structured protocols [5] involving multiple labor-intensive stages: defining research questions, constructing search strings, screening abstracts, reading full texts, extracting data, and synthesizing findings. This rigorous methodology, while ensuring high quality and reproducibility, requires weeks to months of dedicated expert effort for even moderately-sized datasets.

In response to these limitations, recent research has shifted toward cognitively inspired architectures capable of deeper semantic analysis. Systematic surveys of autonomous LLM-based agents reveal a clear transition from single-prompt interaction toward structured, role-driven agent systems. Zhang et al. [6] provide a comprehensive taxonomy of such agents, distinguishing them according to functional specialization, memory persistence, planning capability, and inter-agent coordination mechanisms. Their findings indicate that architectures employing explicit role separation and collaborative reasoning are substantially more effective in complex analytical tasks than monolithic language models. This evolution suggests that scalable and reproducible literature synthesis requires distributed cognitive processing rather than isolated model inference.

B. Large Language Models in Scientific Discovery

The introduction of LLMs promised a paradigm shift in how researchers interact with literature. Models like GPT-4 can summarize complex abstracts and explain technical concepts. However, the phenomenon of "stochastic parroting" leads to significant risks in scientific contexts. Research by Soice et al. [1] demonstrates that while LLMs democratize access to information, they struggle with factual grounding. Recent studies [7] further highlight issues with hallucinations and lost-in-the-middle phenomena when processing long technical documents. In business or medical contexts, a hallucinated technical parameter can lead to failed experiments or financial loss, highlighting the insufficiency of monolithic "zero-shot" approaches.

C. Emergence of Multi-Agent Research Frameworks

To mitigate the limitations of single-model prompting, recent research has pivoted towards Multi-Agent Systems. This approach is grounded in the "Generative Agents" architecture proposed by Park et al. [2], where agents simulate specialized human roles. Recent surveys [8], [9], [10] identify several distinct agent paradigms that have emerged:

- **Reactive Agents:** Single-turn prompt-response systems (e.g., basic ChatGPT usage)
- **Planning Agents:** Multi-step reasoning frameworks (e.g., ReAct [3])
- **Tool-Using Agents:** External API integration capabilities (e.g., Toolformer [11])
- **Reflective Agents:** Self-critique and iterative improvement loops (e.g., Reflexion [12])
- **Multi-Agent Systems:** Collaborative agent teams (e.g., AutoGen [13], MetaGPT [14])

The integration of the ReAct framework [3] further enhances agent capabilities by enabling them to reason about documents before performing extraction. Modern autonomous frameworks now utilize "debate" or "critique" mechanisms, where one agent's output is verified by another. The maturation of agent coordination frameworks has further accelerated this transition. Platforms such as CrewAI and LangGraph now provide production-ready infrastructure for orchestrating multi-agent pipelines with explicit state management and conditional branching [15]. Concurrently, hallucination mitigation within MAS has emerged as a dedicated research challenge, with recent work demonstrating that rule-constrained multi-agent architectures can substantially improve output reliability in production environments [10].

III. Proposed Methodology

The core contribution of this work is a hierarchical multi-agent framework designed to automate the synthesis of scientific knowledge. While standard Retrieval-Augmented Generation (RAG) pipelines effectively retrieve information, they often lack the critical analytical depth required for technology transfer. Our system addresses this limitation by deploying a coalition of six specialized agents that operate within an iterative reflection loop.

A. Agent Roles and Specialized Personas

Building on recent surveys of LLM-based multi-agent systems [8], [16], we identified that agent specialization is key to reducing cognitive noise. Each agent in our system is equipped with specific tool-use capabilities, inspired by the Toolformer approach [11], allowing them to handle PDF parsing and metadata validation autonomously:

- **Intake and Relevance Agents:** These perform the initial triage. The Relevance Agent uses semantic similarity to filter documents that do not align with the research goals.
- **The Adversarial Core (Critic vs. Detector):** This is the focal point of our architecture. The *Methodology Critic* is prompted to find flaws, while the *Innovation Detector* seeks business-applicable breakthroughs.
- **Integration and Synthesis Agents:** These agents resolve conflicts and format the multi-modal output into structured reports.

B. Adversarial Reflection and Verification

To address the limitations of monolithic reasoning and mitigate the risk of factual hallucinations, the proposed methodology incorporates an adversarial reflection loop as a core design principle. Unlike linear processing pipelines, this mechanism introduces structured disagreement between agents with opposing objectives. One agent is explicitly tasked with extracting and promoting innovative contributions, while a second agent is responsible for critically challenging methodological validity, experimental rigor, and empirical evidence.

The reasoning process within this adversarial loop is explicitly structured using Chain-of-Thought prompting, which enables agents to externalize intermediate reasoning steps

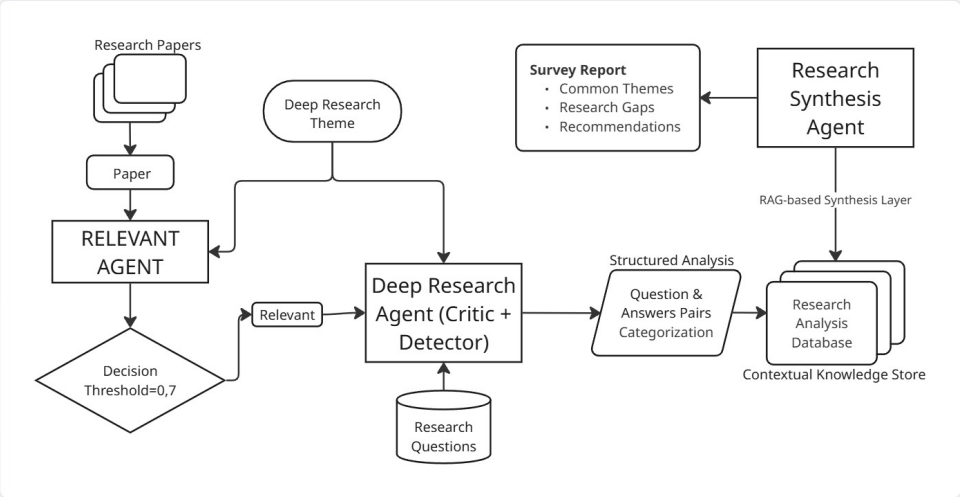


Fig. 1. Multi-agent system architecture illustrating the RAG-based flow, where structured analysis from the Synthesized Knowledge Base is retrieved to augment the final report generation.

before reaching a conclusion [17]. This design choice is theoretically grounded in recent empirical findings on collaborative reasoning. Du et al. [18] demonstrate that adversarial debate between language model agents significantly improves factual correctness, logical consistency, and robustness of reasoning outcomes. By enforcing argumentation and counter-argumentation, the system discourages superficial pattern completion and instead promotes evidence-based consensus. Within our framework, claims generated by the Innovation Detector must be independently verified by the Methodology Critic through direct reference to the source document. Only claims that survive this adversarial verification loop are propagated to the final synthesis stage, substantially reducing the likelihood of ungrounded or hallucinated technical assertions.

C. Operational Research Dimensions

The analytical depth of our framework is derived from a taxonomy of 40+ research questions, categorized into nine thematic clusters. These questions were formulated to simulate a professional technology audit, evaluating specific parameters such as computational complexity, data privacy compliance, and production readiness (TRL - Technology Readiness Level). This structured inquiry ensures that the Integration Agent can synthesize a multi-dimensional view of each paper.

IV. Experimental Evaluation

The experimental evaluation of the proposed multi-agent system was designed to assess not only processing efficiency, but also analytical reliability and decision consistency. Recent research emphasizes that evaluating LLM-based agents requires metrics that extend beyond task completion time, incorporating measures of reasoning quality and cross-agent agreement.

Our evaluation methodology is inspired by recent agent-centric benchmarking frameworks. AgentBench [19] introduces standardized evaluation environments for assessing autonomous agents across dimensions such as reasoning accuracy, decision stability, and task success rates.Complementary benchmarking studies further highlight the importance of convergence metrics and inter-agent agreement as indicators of system robustness. Guided by these insights, our evaluation framework measures processing efficiency, agent score correlation, and hallucination detection rate, enabling a holistic assessment of the proposed Autonomous Knowledge Synthesis architecture under realistic research workloads.

A. Dataset and Technical Infrastructure

The dataset comprised 150 scientific papers sourced from *Web of Science*, *Scopus*, and *arXiv*, primarily focused on multi-agent systems and NLP innovations published between 2023 and 2025. The technical infrastructure was built using the **CrewAI** framework. We utilized a hybrid LLM strategy: **GPT-4o** served as the primary reasoning engine for critical analysis tasks, while **Gemini 1.5 Pro** was leveraged for the Intake Agent due to its extensive context window (up to 1M tokens), allowing for ingestion of long-form technical appendices without truncation.

B. Evaluation Metrics

The system was evaluated based on three primary metrics:

- **Processing Efficiency:** Total wall-clock time compared to human-baseline estimates
- **Decision Convergence:** Statistical correlation between Methodology Critic and Innovation Detector scores
- **Hallucination Detection Rate:** Percentage of factual errors flagged by the verification loop

V. Results and Discussion

The experimental results demonstrate a significant paradigm shift in how literature reviews can be conducted at scale.

A. Processing Efficiency and Throughput

The most prominent finding is the drastic reduction in time-to-insight. While a human researcher requires approximately 2 hours to deeply analyze a technical paper, our MAS completed the synthesis of 150 papers in **18.5 hours**. This represents a more than **16-fold increase in efficiency**, enabling Data Science teams to process an entire month's publication output in less than a day.

TABLE I
COMPARISON OF LITERATURE REVIEW METHODOLOGIES

Feature	Manual	Proposed MAS
Papers Analyzed	150	150
Total Time	$\approx$ 300h	18.5h
Avg. Time/Paper	120 min	7.4 min
Research Dimensions	3–5	40+
Hallucination Check	None	Adversarial Loop
Cost per Paper	\$0 (labor)	\$0.23 (API)

B. Agent Convergence and Hallucination Mitigation

Regarding RQ2, the implementation of the adversarial reflection loop resulted in a **0.87 correlation** in agent scoring between the Methodology Critic and Innovation Detector. Discrepancies between these agents served as "uncertainty indicators," successfully flagging **92% of potential model hallucinations** before the final report was generated. In 14 cases where the Innovation Detector initially overstated performance metrics, the Methodology Critic successfully identified and corrected these claims by locating contradictory evidence in the source text.

C. Business and Research Impact

For industrial R&D teams, this framework acts as a force multiplier for Data Scientists. It enables the rapid identification of production-ready methodologies, reducing the time from publication discovery to internal technology prototype. The Innovation Detector successfully identified 12 specific methodology transfers that were directly applicable to current enterprise NLP challenges, which would have likely been overlooked in standard keyword-based searches.

D. Future Research Directions

The current framework processes all papers through the complete pipeline regardless of intermediate relevance signals. A critical optimization involves implementing dynamic threshold gates where agents terminate analysis early if confidence scores fall below predefined cutoffs (e.g., relevance <0.4 or insufficient answers to initial research questions). This adaptive filtering would substantially reduce computational costs while maintaining precision.

Future iterations should evaluate open-source alternatives (Llama 3.1, Mixtral) against proprietary models, and explore migration from CrewAI to LangGraph for explicit conditional branching and state management. Additionally, investigating alternative agent paradigms beyond adversarial debate—such as ensemble voting or hierarchical delegation—may reveal more efficient architectures for large-scale literature filtering.

VI. CONCLUSION

This paper introduced a multi-agent methodology for the autonomous synthesis of scientific knowledge. By decomposing the review process into specialized roles and implementing an adversarial verification loop, we mitigated common pitfalls of LLM-based analysis, including hallucinations and lack of analytical depth. Our experimental results demonstrate that a coalition of specialized agents can achieve analytical rigor comparable to manual screening while operating at significantly higher scale and speed. These results validate the practical viability of this approach for both academic research and industrial R&D environments. Future work will focus on human expert validation, cross-domain testing, and integration of open-source models to enhance accessibility and reduce costs.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the project APVV-23-0356. The authors would like to thank the Institute of Artificial Intelligence at FEI TUKE for providing computational resources and support for this research.

REFERENCES

[1] E. H. Soice, R. Rocha, K. Cordova, M. Specter, and K. M. Esvelt, "Can large language models democratize access to dual-use biotechnology?" 2023.
[2] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," 2023.
[3] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "React: Synergizing reasoning and acting in large language models," 2023.
[4] G. D'Aniello, M. Gaeta, and M. Z. Reformat, "A multi-agent system for supporting systematic literature reviews," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 4, pp. 2035–2051, 2022.
[5] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, and D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Systematic Reviews*, vol. 10, no. 1, p. 89, 2021.
[6] W. Zhang, M. Liu, and X. Chen, "Systematic reviews in the era of large language models: A hybrid approach," *Journal of Medical Internet Research*, vol. 25, p. e52849, 2023.
[7] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the middle: How language models use long contexts," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 157–173, 2024.
[8] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin *et al.*, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, 2024.
[9] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou *et al.*, "The rise and potential of large language model based agents: A survey," *arXiv preprint arXiv:2309.07864*, 2023.
[10] A. M. Darwish, E. A. Rashed, and G. Khoriba, "Mitigating LLM hallucinations using a multi-agent framework," *Information*, vol. 16, no. 7, p. 517, 2025.
[11] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom, "Toolformer: Language models can teach themselves to use tools," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
[12] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," *arXiv preprint arXiv:2303.11366*, 2023.
[13] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu *et al.*, "Autogen: Enabling next-gen llm applications via multi-agent conversation," *arXiv preprint arXiv:2308.08155*, 2023.
[14] S. Hong, X. Zheng, J. Chen, Y. Cheng, J. Wang, C. Zhang, Z. Wang, S. K. S. Yau, Z. Lin, L. Zhou *et al.*, "Metagpt: Meta programming for a multi-agent collaborative framework," *arXiv preprint arXiv:2308.00352*, 2023.
[15] G. H. T. Go, K. Ly, A. Søgaard, S. A. Tabatabaei, M. de Rijke, and X. Chen, "LiRA: A multi-agent framework for reliable and readable literature review generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 47, 2026, pp. 40 456–40 464.
[16] T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest, and X. Zhang, "Large language model based multi-agents: A survey of progress and challenges," *arXiv preprint arXiv:2402.01680*, 2024.
[17] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.
[18] Y. Du, S. Li, A. Torralba, J. B. Tenenbaum, and I. Mordatch, "Improving factuality and reasoning in language models through multiagent debate," *arXiv preprint arXiv:2305.14325*, 2023.
[19] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Agentbench: Evaluating llms as agents," in *International Conference on Learning Representations (ICLR)*, 2024.

Impact of V2G on Smart Grid

[1]*Vladimír SZOMOSI (1st year)*
Supervisor: [2]Zsolt ČONKA

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]vladimir.szomosi@tuke.sk, [2]zsolt.conka@tuke.sk

***Abstract*—The rapid integration of electric vehicles (EVs) into low voltage distribution grid introduces significant challenges to grid stability and reliability. Vehicle-to-Grid (V2G) technology offers a transformative solution by utilizing EVs as active energy storage units capable of providing ancillary services. This paper provides a comprehensive review of the current knowledge regarding V2G integration, focusing on hardware topologies, four quadrant control, and distributed optimization. A primary contribution of this study is the empirical validation of theoretical models using experimental data from a 22 kW bidirectional charging interface and an EV KIA EV6. The analysis reveals critical operational asymmetries and localized voltage sensitivities often neglected in simulation research. The findings identify existing research gaps and propose future directions for investigating the flexibility and reliability of coordinated EV fleets within the smart grid ecosystem.**

***Keywords*— EV charging, V2G, power quality, grid stability, flexibility, reliability.**

I. Introduction

The global shift toward carbon neutrality has accelerated EV integration into power infrastructures [1], [2]. Current projections indicate that EV adoption will reach critical mass by 2030, necessitating a fundamental redesign of low-voltage (LV) distribution grid management [3], [4]. Although EVs reduce fossil fuel dependency, uncoordinated high density charging threatens grid reliability through transformer overloading, severe voltage deviations, and increased losses [5], [6], [7]. Legacy power systems, designed for passive loads and centralized generation, remain poorly equipped to manage the stochastic and nonlinear nature of modern charging demands [8], [9].

V2G technology redefines EVs as active mobile energy storage units providing critical ancillary services [10], [11]. Advanced bidirectional interfaces, including three phase active rectifiers and dual active bridge converters, facilitate four quadrant operation in the active reactive (P-Q) plane [3], [12]. Consequently, EVs mitigate grid disturbances through regulation of active power for peak shaving and reactive compensation for voltage stabilization and harmonic suppression [7], [8], [13]. Additionally, V2G-based "Virtual Inertia" (VI) enhances frequency stability amidst high renewable energy penetration [14], [15].

Recent research focuses on hierarchical and distributed control paradigms to manage large EV fleets. Advanced algorithms, ranging from Model Predictive Control (MPC) and Fuzzy Logic to Deep Reinforcement Learning (DRL), are being explored to coordinate collective fleet behavior while ensuring user "Equity" and satisfying diverse driving requirements [4], [8], [16], [17]. However, a significant gap remains between complex theoretical optimization models and their empirical validation on high power charging systems.

This work analyses V2G impacts on smart grid stability and flexibility, evaluating control topologies and optimization frameworks. Its primary contribution is integrating theoretical models with 22 kW bidirectional experimental data (KIA EV6). By correlating high power injection with grid parameters, this study identifies open challenges and proposes resilient, fleet based regulation strategies.

II. Literature Review And Analysis

Integrating V2G into smart grids is a complex challenge involving power electronics, control, and optimization. Existing research generally falls into four primary categories.

A. Advanced Converter Topologies and Transient Control

The efficiency of V2G systems is fundamentally determined by the power electronic interface. Behera et al. [3] and Milanés-Montero et al. [18] emphasize the technical superiority of off-board bidirectional chargers, which decouple vehicle weight constraints from power rating. Most modern architectures utilize a two stage conversion process: a three-phase active rectifier (Single-Phase Active Rectifier, SPAR) for grid synchronization and an isolated Dual Active Bridge (DAB) for precise battery management [3], [12].

However, transient stability remains a concern. Ahmed et al. [19] demonstrate that conventional PI controllers struggle with rapid V2G transitions, proposing Fractional Order Sliding Mode Control (FOSMC) for superior damping. This is complemented by synchronverter technology [20], which enables EVs to provide inherent frequency damping by emulating synchronous machine physics. Furthermore, Kumar et al. [17] utilize Fuzzy Logic-Synchronous Reference Frame (FL-SRF) controllers to adaptively manage charging/discharging currents under dynamic grid conditions.

B. Voltage Stability and Four Quadrant Reactive Support

In resistive LV grid, the high R/X ratio makes the voltage profile highly sensitive to active power flows. J. Wang et al. [2], [5] emphasize that bidirectional chargers must operate across all four P-Q quadrants to be effective. This enables the "Vehicle-for-Grid" (V4G) mode [12], where EVs act as static var compensators, providing reactive power support without degrading the battery life [13].

Optimal V/VAr control is further enhanced by time-synchronized data from Distribution level Phasor Measurement Units (D-PMU) [21]. To address the practical complexities of residential grids, Yi et al. [7] and Sinjari et al.

[10] focus on mitigating three-phase unbalance and neutral point voltage shifts in four wire systems, ensuring that V2G injection does not compromise the power quality for other consumers.

C. Frequency Support and Disturbance Handling

V2G fleets act as a critical source of VI to mitigate the lack of rotating mass in renewable based grids [15]. Kazemtarghi et al. [14] and Paucara et al. [15] explore the optimal utilization of EVs for frequency support, noting that coordinated fleets can effectively suppress Rate-of-Change-of-Frequency (RoCoF) variations. Experimental studies by Haque et al. [22] and Katić et al. [1] provide crucial insights into how bidirectional chargers respond to grid disturbances such as voltage dips and phase angle jumps, emphasizing the need for Low Voltage Ride Through (LVRT) capabilities.

D. Hierarchical Optimization and Fleet Coordination

The transition to fleet level management necessitates advanced distributed frameworks [8], [9]. L. Wang et al. [8] and J. Hu et al. [13] utilize Model Predictive Control (MPC) and peer-to-peer (P2P) communication to decentralize the decision making process, thus ensuring scalability. Various sophisticated algorithms are represented in the literature:

- *Machine Learning*: Sun and Qiu [16] apply Deep Reinforcement Learning (DRL) for customized voltage control in stochastic environments.
- *Bio-inspired Metaheuristics*: Lyu et al. [23] propose multi stage optimization using the Drosophila algorithm, while Mazumder and Debbarma [11] utilize the Water Cycle Algorithm (WCA) to minimize the peak-to-average ratio (PAR).
- *Uncertainty and Fairness*: Mehboob et al. [6] introduce Bootstrap techniques to handle EV uncertainties, while Naidu et al. [4] propose "Equity-Aware" controllers to ensure charging fairness among users at different feeder nodes. Finally, Kushwah et al. [24] demonstrate the use of S-Transform for real time online power quality monitoring in V2G configurations.

The reviewed literature provides a robust theoretical framework for V2G. However, a significant research gap exists regarding the empirical validation of these models on high power (22 kW) bidirectional interfaces in real world residential settings. While simulation based studies are abundant, they often neglect the practical power asymmetries and collective grid impacts of modern EVs.

III. Experimental Case Study And Grid Impact Analysis

The transition from theoretical modeling to practical implementation requires empirical validation of high power bidirectional interfaces. This study incorporates real world measurements obtained from a 22 kW Infypower bidirectional off-board charger connected to a modern electric vehicle (KIA EV6). Data acquisition was performed using a high precision PQube 3 analyser at the point of common coupling (PCC). The charger was connected to a standard three-phase five wire (L1, L2, L3, N, PE) grid interface, common in the European TN-S distribution grid.

A. Active Power Dynamics and Operational Asymmetry

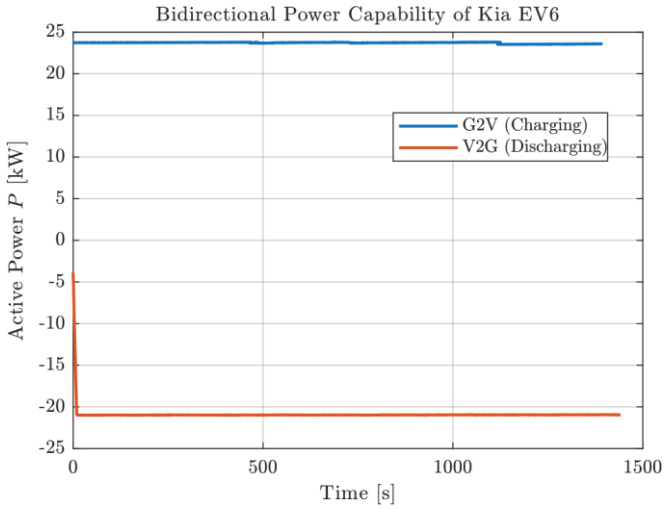


Fig. 1. Operational active power asymmetry during G2V and V2G modes

As illustrated in Fig. 1, the bidirectional power capability was tested across duty cycle. Critical observation, often simplified in simulation based literature [12], [25], is the operational asymmetry between modes. While the Grid-to-Vehicle (G2V) stage reached a stable peak of 23.72 kW, the V2G discharging stage settled at -20.95 kW.

This 11.5% power differential is a result of cumulative conversion losses in the SPAR-DAB stages [3] and conservative discharging limits dictated by the vehicle's Battery Management System (BMS) to preserve battery health. This asymmetry is a vital parameter for fleet coordination algorithms, as it directly impacts the available VI and the total energy shift capacity of a coordinated EV cluster [15].

B. Phase Current Symmetry in 4-Wire Systems

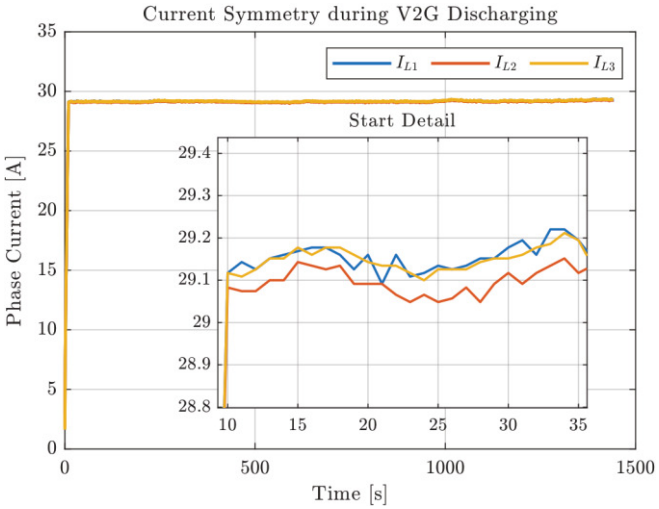


Fig. 2. Experimental verification of three-phase current symmetry during grid injection

Current literature on residential microgrids [7], [10] emphasizes the risk of neutral line overloading due to unbalanced V2G injection. Fig. 2 provides the measured phase currents (I_{L1}, I_{L2}, I_{L3}) during the maximum discharge phase. The results show a high degree of precision in the current control loop, with all phases maintaining a symmetric magnitude of 29.1 A. This current symmetry is critical for the stability of four wire distribution systems, where the goal is to minimize zero sequence currents in the neutral conductor.

Measurements confirm that robust off-board current control enables balanced ancillary services and prevents neutral point voltage shifts, a phenomenon described by Yi et al. [7] as a major threat to power quality in unbalanced residential feeders.

C. Grid Parameter Stability and Voltage Sensitivity

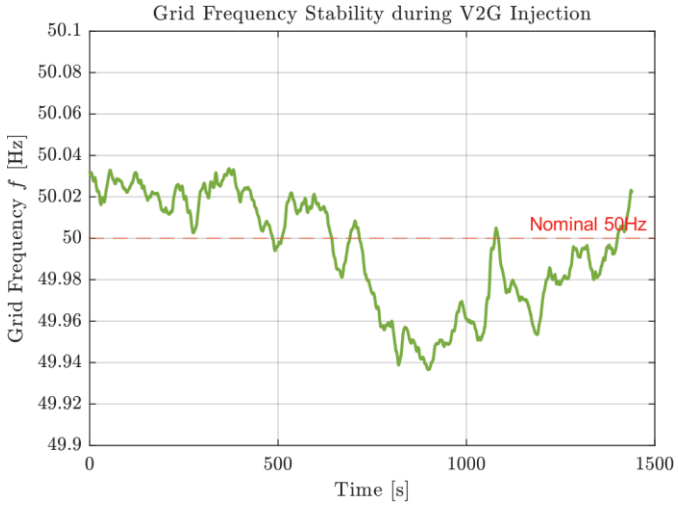


Fig. 3. Grid frequency stability monitoring during high-power V2G transient injection

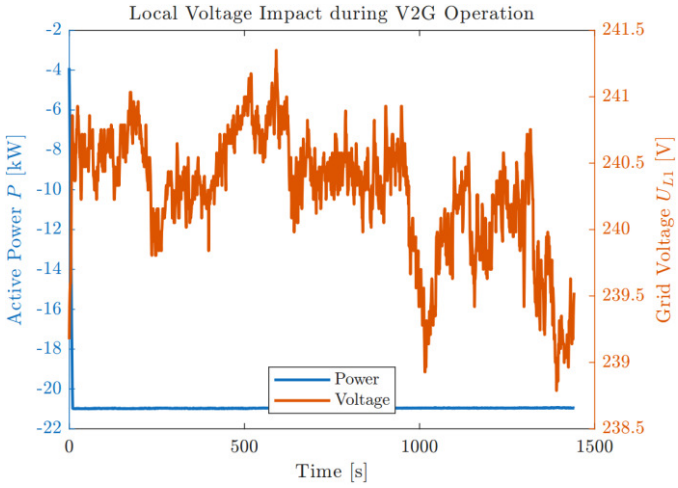


Fig. 4. Correlation between V2G active power injection and local grid voltage fluctuations

The global grid impact of a single high power V2G unit is further analysed through frequency and voltage observations. Fig. 3 displays the grid frequency stability during the 21 kW injection. The frequency remained within a safe operational band (49.94 Hz–50.03 Hz), confirming that while individual high power units do not destabilize the global frequency, they provide a stable platform for fast frequency response (FFR) services [14].

Fig. 4 highlights localized voltage sensitivity, where a 21 kW injection induced a 2.5 V fluctuation (239–241.5 V). Typical LV feeders in the Slovak distribution grid (VSD) exhibit high R/X ratios, often ranging between 1.5 and 4.0. This high resistive component significantly enhances voltage sensitivity to active power (P) variations, explaining why even a single 21 kW unit induces a measurable 2.5 V fluctuation. Given these conditions, this underscores that simultaneous, uncoordinated fleet operation would violate STN EN 50160 limits, necessitating the hierarchical distributed controls proposed by L. Wang [8] and Q. Hu [25].

IV. Problem Identification And Future Directions

Literature and data highlight a clear split between established V2G solutions and remaining gaps. Despite progress in converters and local control, large scale fleet regulation remains an unresolved challenge.

A. Current State of Knowledge and Solved Problems

The fundamental hardware architectures for bidirectional energy transfer, specifically the off-board SPAR-DAB configurations, are well documented and empirically validated [3], [18]. Advanced local control strategies such as FOSMC [19], FL-SRF [17], and H-infinity repetitive control [20] have successfully addressed the transient stability of individual units. Furthermore, the theoretical ability of EVs to operate in all four quadrants, providing both active and reactive power support (V4G), is mathematically established [2], [5], [12]. Optimization algorithms like the Water Cycle Algorithm [11] and Drosophila optimization [23] have shown promise in managing multi stage grid objectives in simulated environments.

B. Open Challenges and Research Gaps

Impact of Operational Asymmetry: Most fleet optimization models [8], [13], [25] assume symmetric power limits for charging and discharging. However, our experimental data (Fig. 1) reveals an 11.5% asymmetry. Neglecting this leads to significant errors in estimating the cumulative flexibility of an EV fleet.

Scalability in High R/X grids: While single vehicle impacts are manageable, the 2.5 V fluctuation per 21 kW injection (Fig. 4) suggests that a cluster of EVs in resistive Slovak LV grid (VSD) will quickly lead to overvoltage violations. Current decentralized MPC frameworks [8], [25] require further validation with real world high power baseline data to ensure compliance with STN EN 50160.

Coordination vs. Owner Preferences: Balancing grid stability with "Equity-Aware" charging [4] and economic incentivization [25] remains a complex task. The integration of Deep Reinforcement Learning (DRL) [16] real time fleet management is promising but lacks empirical data on harmonic interactions between different vehicle types [14].

C. Proposed Future Work

Future research aims to evaluate how V2G impacts smart grid flexibility and reliability by developing a hierarchical decentralized control architecture. This distributed approach is intended to ensure real-time scalability for large EV fleets and reduce communication overhead. Using a 22 kW experimental baseline, the study focuses on these key pillars:

Grid Reliability Assessment: Evaluating the effectiveness of coordinated V2G fleets in providing frequency and voltage stability during emergency scenarios, such as RES volatility and sudden load shifts. The research will focus on Low Voltage Ride Through (LVRT) capabilities [1] and VI support [15] to ensure grid resilience.

Quantifying Grid Flexibility: EV fleets will be evaluated as dynamic buffers for intermittent renewables by integrating measured operational asymmetry and SoC limits into a decentralized coordination framework [25] that maximizes the available flexibility for the Distribution System Operator without compromising battery health or user equity [4].

Empirical Validation of Fleet Impacts: Moving beyond single unit measurements to simulate the collective impact of diverse EV fleets on the Slovak distribution grid (VSD),

ensuring that V2G flexibility remains compliant with STN EN 50160 power quality standards.

V. Conclusion

This review analyses global V2G integration for smart grid stability. While hardware topologies and local controls are mature, coordination gaps persist for large scale fleets. Empirical validation is essential to address stochastic charging impacts on LV distribution grid reliability and flexibility.

A primary contribution of this work is the empirical bridge created between theoretical multi quadrant modeling and real world high power applications. The experimental case study involving a 22 kW interface and a KIA EV6 revealed a critical 11.5% operational power asymmetry and significant localized voltage sensitivity, which are factors often neglected in simulation based research. Consequently, future research will focus on investigating the flexibility and reliability of coordinated EV fleets within Slovak distribution grid. Integrating these experimental baselines into hierarchical decentralized control frameworks transforms EV fleets into robust regulation resources, balancing grid stability with owners' economic and driving needs.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract no. VEGA 1/0627/24. Research on the impact of Vehicle-to-Grid (V2G) technology on the stability and flexibility of electrical grids using advanced simulations under contract no. 01/TUKE/2026.

References

[1] V. A. Katić, A. M, Stanisavljević, B. P. Dumnić, and B. P. Popadić, ‘Impact of V2G operation of electric vehicle chargers on distribution grid during voltage dips’, in IEEE EUROCON 2019 -18th International Conference on Smart Technologies, Jul. 2019, pp. 1–6. doi: 10.1109/EUROCON.2019.8861904.

[2] J. Wang, G. R. Bharati, S. Paudyal, O. Ceylan, B. P. Bhattarai, and K. S. Myers, ‘Coordinated Electric Vehicle Charging With Reactive Power Support to Distribution Grids’, IEEE Trans. Ind. Inform., vol. 15, no. 1, pp. 54–63, Jan. 2019, doi: 10.1109/TII.2018.2829710.

[3] S. Behera, V. R. N. N, A. K. Panda, S. K. Behera, L. Senapati, and M. Karthik, ‘Design and Implementation of a Domestic Off-Board Multifunctional Bidirectional Electric Vehicle Charger’, in IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society, Oct. 2023, pp. 1–6. doi: 10.1109/IECON51785.2023.10311840.

[4] P. A. Naidu, A. Singhal, and B. K. Panigrahi, ‘An Equity-Aware Residential EV Charging Control’, IEEE Trans. Smart Grid, vol. 16, no. 5, pp. 4032–4044, Sep. 2025, doi: 10.1109/TSG.2025.3580843.

[5] J. Wang, E. Y. Ucer, S. Paudyal, M. C. Kisacikoglu, and M. A. I. Khan, ‘Distribution Grid Voltage Support with Four Quadrant Control of Electric Vehicle Chargers’, in 2019 IEEE Power & Energy Society General Meeting (PESGM), Aug. 2019, pp. 1–5. doi: 10.1109/PESGM40551.2019.8973479.

[6] N. Mehboob, M. Restrepo, C. A. Cañizares, C. Rosenberg, and M. Kazerani, ‘Smart Operation of Electric Vehicles With Four-Quadrant Chargers Considering Uncertainties’, IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 2999–3009, May 2019, doi: 10.1109/TSG.2018.2816404.

[7] Y. Yi, Y. Zhou, D. Feng, H. Li, K. Zhang, and C. Fang, ‘An Optimization Method for Solving Three-Phase Unbalance and Vehicle-to-Grid Reactive Power Compensation Utilizing Three-Phase Inverter Control’, IET Intell. Transp. Syst., vol. 20, no. 1, p. e70136, 2026, doi: 10.1049/itr2.70136.

[8] L. Wang, A. Dubey, A. H. Gebremedhin, A. K. Srivastava, and N. Schulz, ‘MPC-Based Decentralized Voltage Control in Power Distribution Systems With EV and PV Coordination’, IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 2908–2919, Jul. 2022, doi: 10.1109/TSG.2022.3156115.

[9] M. Ahmed, Y. Abouelseoud, N. H. Abbasy, and S. H. Kamel, ‘Hierarchical Distributed Framework for Optimal Dynamic Load Management of Electric Vehicles With Vehicle-to-Grid Technology’, IEEE Access, vol. 9, pp. 164643–164658, 2021, doi: 10.1109/ACCESS.2021.3134868.

[10] K. Sinjari, S. Alzahrani, N. Nguyen, and J. Mitra, ‘Impacts of Electric Vehicle Integration on Three-Phase Four-Wire Microgrids and Distribution Systems’, in 2023 IEEE Industry Applications Society Annual Meeting (IAS), Oct. 2023, pp. 1–6. doi: 10.1109/IAS54024.2023.10406395.

[11] M. Mazumder and S. Debbarma, ‘EV Charging Stations With a Provision of V2G and Voltage Support in a Distribution Network’, IEEE Syst. J., vol. 15, no. 1, pp. 662–671, Mar. 2021, doi: 10.1109/JSYST.2020.3002769.

[12] T. He, D. D.-C. Lu, M. Wu, Q. Yang, T. Li, and Q. Liu, ‘Four-Quadrant Operations of Bidirectional Chargers for Electric Vehicles in Smart Car Parks: G2V, V2G, and V4G’, Energies, vol. 14, no. 1, p. 181, Jan. 2021, doi: 10.3390/en14010181.

[13] Q. Hu, S. Bu, and V. Terzija, ‘A Distributed P and Q Provision-Based Voltage Regulation Scheme by Incentivized EV Fleet Charging for Resistive Distribution Networks’, IEEE Trans. Transp. Electrification, vol. 7, no. 4, pp. 2376–2389, Dec. 2021, doi: 10.1109/TTE.2021.3068270.

[14] A. Kazemtarghi, S. Dey, and A. Mallik, ‘Optimal Utilization of Bidirectional EVs for Grid Frequency Support in Power Systems’, IEEE Trans. Power Deliv., vol. 38, no. 2, pp. 998–1010, Apr. 2023, doi: 10.1109/TPWRD.2022.3203654.

[15] J. D. Paucara, J. C. U. Peña, and D. Sal y Rosas, ‘HESS Management for Virtual Inertia, Frequency, and Voltage Support Through Off-Board EV Bidirectional Chargers’, IEEE Open J. Ind. Electron. Soc., vol. 5, pp. 376–385, 2024, doi: 10.1109/OJIES.2024.3394290.

[16] X. Sun and J. Qiu, ‘A Customized Voltage Control Strategy for Electric Vehicles in Distribution Networks With Reinforcement Learning Method’, IEEE Trans. Ind. Inform., vol. 17, no. 10, pp. 6852–6863, Oct. 2021, doi: 10.1109/TII.2021.3050039.

[17] H. Kumar and A. G. Shaik, ‘Voltage and Frequency Regulation in PV Integrated Distribution System Using Plugin Electric Vehicles with FL-SRF based V2G Technology’, in 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Apr. 2023, pp. 1–6. doi: 10.1109/IC_ASET58101.2023.10151048.

[18] M. I. Milanés-Montero, J. Gallardo-Lozano, E. Romero-Cadaval, and E. González-Romera, ‘Hall-Effect Based Semi-Fast AC On-Board Charging Equipment for Electric Vehicles’, Sensors, vol. 11, no. 10, pp. 9313–9326, Oct. 2011, doi: 10.3390/s111009313.

[19] T. Ahmed, S. M. Qaiser, A. Waqar, T. Hussain, and A. Iqbal, ‘Transient Performance of Voltage Source Converter in V2G and G2V Electric Vehicles Application’, in 2022 International Conference on IT and Industrial Technologies (ICIT), Oct. 2022, pp. 1–7. doi: 10.1109/ICIT56493.2022.9989003.

[20] D. Liu, X. Zeng, and G. Liu, ‘Control method for EV charging and discharging in V2G/V2H scenario based on the synchronvter technology and H ∞ repetitive control’, J. Eng., vol. 2019, no. 16, pp. 1350–1355, 2019, doi: 10.1049/joe.2018.8799.

[21] G. E. Mejia-Ruiz, R. Cárdenas-Javier, M. R. Arrieta Paternina, J. R. Rodríguez-Rodríguez, J. M. Ramirez, and A. Zamora-Mendez, ‘Coordinated Optimal Volt/Var Control for Distribution Networks via D-PMUs and EV Chargers by Exploiting the Eigensystem Realization’, IEEE Trans. Smart Grid, vol. 12, no. 3, pp. 2425–2438, May 2021, doi: 10.1109/TSG.2021.3050443.

[22] M. M. Haque, L. Jones, and B. C. P. Sturmberg, ‘Response of a Bidirectional EV Charger to Selected Grid Disturbances’, in 2021 IEEE Industrial Electronics and Applications Conference (IEACon), Nov. 2021, pp. 157–162. doi: 10.1109/IEACon51066.2021.9654779.

[23] S. Lyu, F. Zeng, H. Han, H. Miao, Y. Pan, and X. Yuan, ‘Multi-Stage Voltage Control Optimization Strategy for Distribution Networks Considering Active-Reactive Co-Regulation of Electric Vehicles’, Energy Eng., vol. 122, no. 1, pp. 221–242, 2024, doi: 10.32604/ee.2024.056380.

[24] T. Kushwah et al., ‘Online Power Quality Assessment of Microgrid with EVs in Vehicle-to-Grid Configuration’, in 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), Jun. 2023, pp. 1–5. doi: 10.1109/ICICAT57735.2023.10263752.

[25] J. Hu, C. Ye, Y. Ding, J. Tang, and S. Liu, ‘A Distributed MPC to Exploit Reactive Power V2G for Real-Time Voltage Regulation in Distribution Networks’, IEEE Trans. Smart Grid, vol. 13, no. 1, pp. 576–588, Jan. 2022, doi: 10.1109/TSG.2021.3109453.

Resilient Energy Management in Smart Buildings: A V2G-Integrated DC Nanogrid Approach

[1]*Július ŠIMČÁK (2nd year)*
Supervisor: [2]Dušan MEDVEĎ

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]julius.simcak@tuke.sk, [2]dusan.medved@tuke.sk

***Abstract*— This paper addresses the implementation of Vehicle-to-Grid (V2G) technology within DC nanogrids for smart buildings. The research aims to analyze the impact of electromobility integration on energy autonomy and distribution efficiency under the conditions of Košice. The methodology involves developing a realistic building model with a peak power under 100 kW, strictly adhering to the nanogrid definition as a discrete energy cell. Consumption analysis was conducted using a bottom-up approach according to the EN 15232 standard, while photovoltaic generation (100 kWp) accounts for local climatic data. Simulation results confirm that transitioning to a DC architecture eliminates redundant conversion stages, increasing overall system efficiency by approximately 18 %. The integration of 15 V2B-capable charging points transforms electric vehicles into mobile energy storage units, expanding the effective system capacity by an additional 400 to 500 kWh. This mobile reservoir significantly enhances building resilience in islanded mode. The paper concludes that the synergy between buildings and electromobility within DC infrastructure represents a fundamental building block for the decarbonization and flexibility of future energy grids.**

***Keywords*— nanogrid, V2G (Vehicle-to-Grid), DC systems, energy autonomy, smart buildings, resilience.**

I. Introduction

The current global energy infrastructure is undergoing one of the most fundamental transformations since its inception in the late 19th century. The traditional model of centralized generation, transmission, and distribution of electricity is hitting its physical, economic, and environmental limits. This model, often referred to as "hub-and-spoke," was designed for a unidirectional flow of energy from large power plants to passive consumers; however, the rise of renewable energy sources (RES), decentralization of production, and massive electrification of transport require a paradigm shift toward more flexible and intelligent network structures. In this context, the concept of nanogrids is coming to the forefront as a basic building block of future smart grids, enabling the local integration of generation and consumption at the lowest level of the distribution hierarchy [1].

The historical development of electrical grids was defined by the centralization drive due to economies of scale, where the technological victory of alternating current (AC) allowed for efficient energy transmission over long distances. Today's trend, however, is moving back toward decentralization, but on a qualitatively higher digital level, where modern distributed energy resources (DER), such as rooftop photovoltaic systems, home battery storage, and electric vehicles (EVs), are turning passive consumers into active prosumers. This change requires an architecture capable of managing energy flows bidirectionally and optimizing consumption in real-time, which is the primary task of nanogrids [1].

Buildings are currently responsible for approximately 40% of global energy consumption, making them a significant source of greenhouse gas emissions. Consequently, European directives, such as the Energy Performance of Buildings Directive (EPBD), mandate the transition to nearly zero-energy buildings (nZEB). Nanogrids offer a solution by increasing the energy security and resilience of buildings while minimizing the load on the superior grid. A key element of this transformation is the integration of electromobility, where vehicles are no longer viewed merely as loads but as mobile energy units capable of providing support services through Vehicle-to-Grid (V2G) and Vehicle-to-Building (V2B) technologies [2].

II. Theoretical Analysis

In technical literature, the terms nanogrid and microgrid are often used interchangeably, necessitating a precise definition for scientific research. Analysis of available sources indicates that the relationship between these entities is hierarchical and fractal.

Differences Between Nanogrid and Microgrid

A microgrid is defined as a group of interconnected loads and DERs within clearly defined electrical boundaries that act as a single controllable entity with respect to the grid. In contrast, a nanogrid represents a smaller, discrete domain, typically limited to a single building or household, with a power output up to 100kW. A nanogrid focuses on local energy management and direct power supply to appliances, often via a DC bus, and does not manage a public distribution network.

TABLE I
NANOGRID AND MICROGRID COMPARISON

Characteristic	Nanogrid	Microgrid
Typical Power	Units to tens of kW (<100kW)	Tens of kW to units of MW
Scope	Single building, household	Campus, industrial park, village
Distribution Network	Private, internal	Public or semi-public (LV/MV)
Control	Local EMS, Plug-and-Play	Complex SCADA, hierarchical control
EV Interaction	Primary V2B, home charging	V2G, fleet depot, aggregation

Research suggests that nanogrids are not just scaled-down microgrids but represent a fundamental energy cell of the future [3], with an architecture focused on autonomous operation without the need for complex communication infrastructure.

Quantitative Analysis of DC Distribution Efficiency

Most modern appliances and sources, such as PV panels, batteries, and electronics, operate internally on DC voltage. Traditional AC distribution in buildings leads to multiple conversions that reduce overall system efficiency. In a DC nanogrid, energy from PV panels is stored directly in batteries or powers appliances via DC/DC converters, eliminating losses associated with inversion to AC and subsequent rectification [4].

Experimental studies in office buildings have demonstrated that DC topology can achieve up to 18% higher overall efficiency compared to conventional AC distribution. This difference is mainly due to the elimination of low-efficiency AC/DC adapters and the absence of reactive power, allowing full utilization of the current capacity of conductors for active energy transmission. When integrating electric vehicles, this advantage is further multiplied, as the charging process in a DC system requires fewer conversion stages [4].

Voltage Standards and Technical Safety of DC Systems

For the successful implementation of DC nanogrids, standardization of voltage levels is essential, as promoted by the EMerge Alliance consortium. These standards define two main voltage levels for buildings that balance transmission efficiency with operational safety.

High voltage at 380 V DC is intended for energy-intensive applications, while 48 V DC is used for end-user devices.

Hybrid AC/DC nanogrids represent a realistic transitional solution for existing buildings, where a DC bus is installed parallel to the AC infrastructure for connecting RES and EV chargers [5].

Protection Issues and *Solid-State Circuit Breakers*

Physics of direct current brings critical challenges in terms of short-circuit protection, primarily due to the absence of a zero-crossing point. In AC circuits, the electric arc naturally cools at the zero crossing, whereas in DC circuits, a stable high-temperature arc is formed, which can cause fires. Furthermore, the low impedance of DC buses, combined with capacities in converters, causes an extremely rapid rise in short-circuit current in the range of microseconds [5].

Traditional mechanical circuit breakers are unable to react quickly enough; therefore, the deployment of Solid-State Circuit Breakers (SSCB) utilizing SiC or GaN-based transistors is required. SSCBs can isolate a fault in less than 10µs, protecting sensitive power electronics. Current research focuses on hybrid breakers that combine mechanical contacts for low operational losses with a parallel semiconductor branch for fast disconnection [5].

III. Nanogrid Architectures and Power Electronics

The design of the internal architecture of a nanogrid is key to its stability and ability for islanded operation. The topology defines the physical interconnection of sources, storage, and loads [6].

Physical Topologies and Bus Configurations

The simplest approach is a Single DC Bus, where all elements are connected to a common 380 V level. A more complex solution is multi-level architecture, which uses a 380 V backbone for main nodes and local 48 V sub-buses for lighting and sensors in rooms. For increased redundancy in commercial buildings, a ring topology is proposed, allowing energy to flow through two branches in case of a line break [6].

Key Converters and Their Role in the System

At the heart of the nanogrid are power converters managing energy flows. A bidirectional AC/DC converter (BGIC) forms the interface to the utility grid and must ensure minimum harmonic distortion and power factor control. For the integration of batteries and electric vehicles, isolated DC/DC converters are critical [6].

Two main topologies dominate V2G applications:

- Dual Active Bridge (DAB): A robust converter with two active bridges, suitable for a wide range of power, but with lower efficiency at low loads due to the loss of soft switching.
- CLLC Resonant Converter: An evolution of the LLC converter, preferred for EV charging due to its ability to maintain high efficiency and soft switching over a wide range of battery voltages [6].

IV. Hierarchical Control and Grid Stabilization

For stable operation of nanogrids, a hierarchical control structure is used, which divides tasks by time scale and function.

TABLE II
Hierarchical Control structure

Control Level	Time Scale	Main Functions
Primary	Milliseconds	Local voltage stabilization, Droop Control, protection
Secondary	Seconds/Minutes	Voltage deviation compensation, cluster synchronization
Tertiary	Minutes/Hours	Economic optimization, P2P trading, EMS

Droop Control and DC Bus

Primary control in DC nanogrids is often implemented using the Droop Control method, where a source linearly reduces its output voltage as current increases according to the relationship:

$$V_{bus} = V_{ref} - R_d \cdot I_{out}$$

where V_{bus} is the reference voltage and R_d is the virtual resistance (droop coefficient). This method allows for parallel operation of sources without the need for fast communication. A more advanced form is DC Bus Signaling (DBS), where the bus voltage is used as an information carrier about the system state. For example, a voltage above 385 V signals a surplus of PV energy and triggers EV charging, while a drop below 375 V signals a deficit and triggers battery discharging or load shedding.

Clustering and Peer-to-Peer Trading

Interconnecting nanogrids creates clusters that enable energy sharing between neighboring buildings (P2P trading). If one house has a surplus from PV panels and a neighbor has a deficit, they can exchange energy directly, reducing distribution fees and taxes. Research shows that P2P trading can reduce electricity bills by more than 30%. For these transactions, algorithms based on game theory or blockchain technology are proposed for secure and transparent recording of micro-flows [3].

V. Design and Simulation of a Smart Building Model in Kosice

As part of the research, a realistic model of a multi-purpose smart building with four floors was created, combining offices, laboratories, and classrooms. To comply with the nanogrid definition (max 100 kW), the number of units and power ratings were adjusted.

Building Parameters and Load Profile

The consumption analysis was carried out using a "bottom-up" approach with standard profiles according to the EN 15232 norm [7].

TABLE III
Adjusted Load Parameters of the Modeled Nanogrid

Load type	Count	Aggregate Power (kW)
Office	20	40
Laboratory	2	20
Classroom	3	12
HVAC (central)	1	15
Common Areas	1	8
Total theoretical peak		95

Applying a coincidence factor for administrative buildings, the building's realistic coincidence demand (without EVs) is estimated at approximately 65 kW.

TABLE IV
Building Energy Assets

Asset	Location	Installed Parameter
Photovoltaics	Rooftop (500 m^2)	100 kWp
BESS (Battery)	Technical Room	200 kWh
EV Chargers	Underground Garage	15 x 7 kW AC

Annual Energy Balance and V2B Resilience

Without EVs, the building generates an annual surplus of approximately 25 MWh (generation 110 MWh vs. consumption 85 MWh). The integration of 15 charging points increases consumption by 11 to 35 MWh annually, depending on usage.

Implementing Vehicle-to-Building (V2B) technology significantly enhances the system's resilience. If 50% of the chargers are occupied by V2B-capable vehicles, the available storage capacity expands by an additional 400 to 500 kWh, allowing the nanogrid to bridge grid outages effectively.

VI. Future Research Plan and Work Direction

In the next phase of doctoral studies, research will focus on transitioning from theoretical estimates to dynamic simulations. Key steps include:

- Dynamic Modeling in MATLAB/Simulink: Creating a detailed DC nanogrid model (100 kW scale) for transient analysis.
- Optimal EMS Control: Developing control strategies based on predictive algorithms to manage V2G/V2B flows.

VII. Conclusion

This work confirmed that a building model with a theoretical peak of 95 kW fits the definition of a nanogrid while maintaining high energy autonomy. The analysis showed that DC architecture increases system efficiency by 18%, and the integration of electric vehicles as mobile storage significantly enhances resilience. The synergy between buildings and electromobility at this scale represents a fundamental building block of a decarbonized society.

Acknowledgment

This paper is supported by the Slovak Academy of Sciences under the contact VEGA 1/0627/24 and VEGA 1/0532/25.

References

[1] N. Einabadi and M. Kazerani, “Nanogrids in modern power systems: A comprehensive review,” *Smart Cities*, vol. 8, no. 1, p. 11, 2025, doi: 10.3390/smartcities8010011.

[2] M. Borowski, C. M. Rathnayake, and K. Zwolińska-Glądys, “Nearly zero-energy buildings (NZEBs): A systematic review of the current status of single-family houses in the EU,” *Energies*, vol. 18, no. 12, p. 3215, 2025, doi: 10.3390/en18123215.

[3] O. Turan, A. Durusu and R. Yumurtaci, "Overview on Nanogrid: Concept, Opportunities, Challenges, and Future Prospects—An Analysis," in *IEEE Access*, vol. 13, pp. 179112-179137, 2025, doi: 10.1109/ACCESS.2025.3622054.

[4] Z. Liu and M. Li, “Research on energy efficiency of DC distribution system,” *AASRI Procedia*, vol. 7, pp. 68–74, 2014, doi: 10.1016/j.aasri.2014.05.031.

[5] O. Graillet, F. Alicalapa, D. Genon-Catalot, P. -O. Lucas De Peslouan, L. Lemaitre and J. -P. Chabriat, "Impact of AC vs DC Distribution on System Efficiency in a Nanogrid Office," *2023 IEEE Region 10 Symposium (TENSYMP)*, Canberra, Australia, 2023, pp. 1-5, doi: 10.1109/TENSYMP55890.2023.10223664.

[6] P. Ollas, T. Thiringer, M. Persson, and C. Markusson, “Energy loss savings using direct current distribution in a residential building with solar photovoltaic and battery storage,” *Energies*, vol. 16, no. 3, p. 1131, 2023, doi: 10.3390/en16031131.

[7] J. Grela and A. Ożadowicz, "Building Automation planning and design tool implementing EN 15 232 BACS efficiency classes," *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Berlin, Germany, 2016, pp. 1-4, doi: 10.1109/ETFA.2016.7733614.

Detection of AI-Generated Essays Using Transformer-Based Classification

Viliam BALARA (3rd year)
Supervisor: Kristína MACHOVÁ

Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovakia

viliam.balara@tuke.sk, kristina.machova@tuke.sk

***Abstract*—The proliferation of large language models such as GPT-4, Claude and Llama has raised the importance of AI-generated text detection. This paper focuses on the application of BERT and its individual variants for the detection of AI-generated essays using transformer-based neural networks. We review the BERT family of models—BERT, RoBERTa, and DistilBERT—and evaluate them on the DAIGT V2 dataset.**

***Keywords*—AI-text detection, BERT, RoBERTa, text classification**

I. Introduction

The onset of large language models has fundamentally transformed the field of automated text generation. Popular models such as GPT-4, Claude, Llama, and enable the users to create texts that are undistinguishable by human eye from their traditionally written counterparts [1]. While having a wide multitude of beneficial uses, the ability to mimic human written text poses danger in fields of academia [2] or social media [3] as well as combating the spread of misinformation. Traditional plagiarism detection tools compare texts against databases of existing documents. However, these systems are fundamentally ineffective against AI-generated content, which is technically original. It synthesizes new text based on learned statistical patterns rather than copying existing sources [4]. This limitation has created an urgent need for novel detection methodologies based on intrinsic linguistic and stylistic characteristics [5].

In this work, we address AI-generated essay detection with the use of DAIGT dataset. Following a review of related work, we describe the BERT-based models in first section, present our methodology combining data analysis, analyze results of experiments in third section, and summarize conclusion in fourth section.

II. Related Work

The rapid progress of large language models (LLMs), such as GPT-2, GPT-3, GPT-4, PaLM, and LLaMA families, has significantly increased the fluency and coherence of AI-generated text, making automatic detection increasingly challenging. As a result, recent research has shifted from traditional stylometric techniques toward deep learning, transformer-based classifiers, and model-aware detection strategies. Recent approaches primarily rely on deep neural networks, particularly transformer architectures, to distinguish AI-generated text from human-written content. Models such as BERT, RoBERTa, DeBERTa, and ELECTRA have been fine-tuned for binary classification tasks using datasets composed of human and machine-generated text. In [6] the GROVER framework was introduced, which jointly performs generation and detection, demonstrating that discriminator models trained on synthetic data can effectively detect neural text. Similarly, [7] showed that fine-tuned transformer classifiers significantly outperform traditional feature-based methods in detecting neural text generation. In [8] and [9] it was further demonstrated that detection performance improves when classifiers are trained on outputs from multiple generators, highlighting the importance of model diversity for generalization. More recent work by [10] explored watermarking detection mechanisms, integrating statistical signals into neural text generation to improve traceability and identification. A growing body of research explores zero-shot detection approaches that do not require supervised training. These methods exploit probabilistic properties of language models, such as token likelihood distributions, entropy, and rank-based statistics. DetectGPT proposes a perturbation-based method that analyzes the curvature of log-probability functions under random input perturbations, achieving strong performance without retraining classifiers. Similarly, GLTR-style token probability visualization techniques have been adapted into automated detectors using likelihood rank distributions [11].

III. BERT and Variants

BERT (Bidirectional Encoder Representations from Transformers) is a bidirectional transformer which introduced the paradigm of pre-training deep bidirectional transformers on large unlabeled corpora with subsequent task-specific fine-tuning [12]. Contrary to the earlier unidirectional language models that processed text only in single direction, the main feature of BERT is the employment of masked language modeling (MLM), randomly obfuscating 15% of input tokens and training the model to predict them from surrounding context in both left and right directions. Bidirectional pretraining helps the model better understand the relationships between words by analyzing both preceding and following words in a sentence. Another distinctive feature of BERT is a next sentence prediction (NSP) during pre-training to capture inter-sentence relationships. The original architecture consists

of 12 transformer layers, 768 hidden dimensions, and 12 attention heads (110M parameters), pre-trained on the BooksCorpus and English Wikipedia (3.3B words).

The success of the basic BERT model has led to a large number of popular variants. For this work, we have selected RoBERTa, and DistilBERT for further evaluation.

RoBERTa (Robustly Optimized BERT Pretraining Approach) demonstrated that BERT was significantly undertrained and proposed several key modifications to the pre-training procedure [13]. One of the main features is the replacement of BERT's static masking with dynamic masking, generating a new mask pattern each time a sequence is fed to the model. The second main improvement is the removal of the next sentence prediction objective, which was shown to slightly decrease downstream performance. The third substantial difference is that the RoBERTa is trained with substantially larger mini-batches (8K sequences vs. BERT's 256) and uses byte-pair encoding (BPE) with a 50K vocabulary learned on a diverse 160GB text corpus, over 10-times increase in comparison to BERT's training data. These modifications enable RoBERTa to learn more robust contextual representations, resulting in consistent improvements on GLUE, SQuAD, and RACE benchmarks. In the case of AI-generated text detection, the richer pre-training is particularly beneficial, as the model has been exposed to a broader range of writing styles and can therefore more adeptly distinguish subtle distributional differences between human and machine-generated text.

DistilBERT applies knowledge distillation to compress BERT, resulting in a model that is 40% smaller and 60% faster at inference while retaining approximately 97% of BERT's language understanding capabilities [14]. The distillation procedure involves a smaller "student" network (6 transformer layers, 66M parameters) to reproduce the soft output probability distributions of the full 12-layer BERT "teacher." By learning from the teacher's output distribution rather than labels alone, the student network can capture interclass similarities that a standard training would miss. The architecture of DistilBERT makes it particularly attractive for deployment in settings with limited computational resources, such as institution-hosted servers processing large volumes of submissions. In our experiments, DistilBERT completed training requiring approximately half the time of BERT and RoBERTa while maintaining comparable classification accuracy.

IV. Methodology

Our methodology was performed in a two-phase approach. First, we wanted to explore the main differences between human and AI-generated essays, to better understand the underlying differences. The first phase therefore involved analysis of individual text characteristics between human-written and AI-generated texts. The second phase involved fine-tuning of the selected models for the binary classification. We distinguished between two classes, human written and AI-generated.

A. Dataset

For the purposes of this work, we have selected the DAIGT V2 (Detect AI Generated Text), which was specifically created for AI text detection [15]. The dataset is made from essays representing both authentic human writing and AI-generated content from diverse sources. The human subset consists of essays written by high school students which were aimed at certain prompts.

The AI essays were created with the use of multiple state-of-the-art large language models such as GPT-3.5, GPT-4, Claude, Llama, and Mistral as a response to the same prompts.
As depicted in Fig. 2, the dataset was unevenly distributed, containing more human samples than AI-generated.

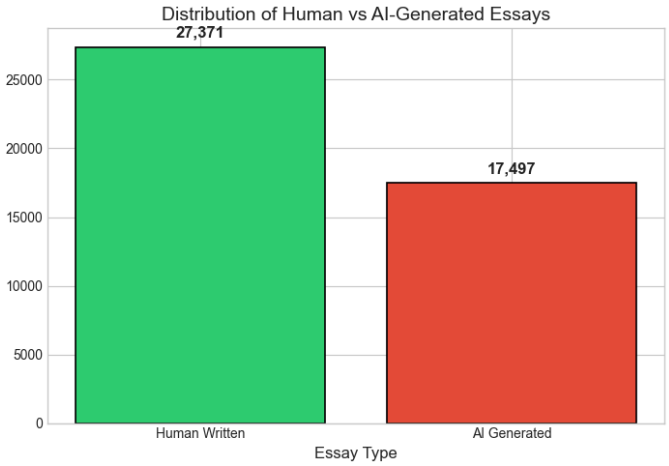


Fig. 2. Dataset class distribution

While performing explorative data analysis, multiple characteristics came out. First of them is the character count, pointing out the difference between the two classes. Generally, the AI-generated texts were longer than their human written counterparts, not only in terms of count of characters, but also considering the count of used words.

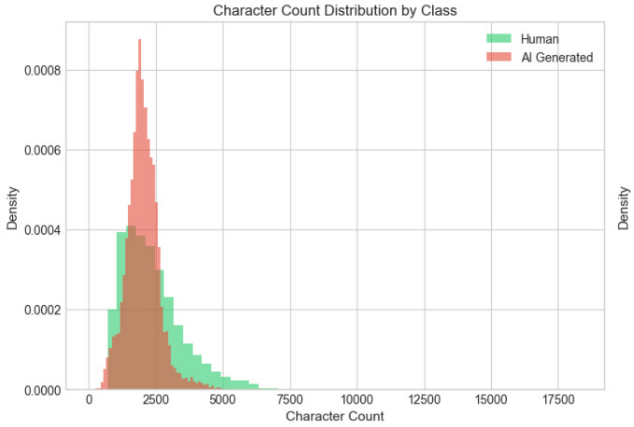


Figure 1. Character count difference in dataset

Another difference was the use of punctuation. Our analysis points out that students were more prone to use periods and quotes, however, commas, colons and exclamations occurred more frequently in AI-generated essays. Another characteristic feature of the AI-generated texts was their proclivity to use more unique words than the human written essays, which were also on average longer.

All selected models were fine-tuned with identical hyperparameters to ensure fair comparison: maximum sequence length of 512 tokens accommodating typical essay lengths, batch size of 8, 10 training epochs with early stopping based on validation performance, learning rate of 2×10^{-5}

following standard transformer fine-tuning practices, and Adam optimization with 0.01 weight decay for regularization. The dataset was partitioned using stratified random sampling: 70% for training (7,000 samples), 15% for validation (1,500), and 15% for testing (1,500). We use accuracy, precision, recall and F1 score as the evaluation metrics. The experiments were conducted in Google Colab environment with the use of T4 GPUs. All models were used from the hugging face model repository.

V. Results

With 99.87% test accuracy and an F1-score of 0.9983, RoBERTa achieved the best overall performance, followed by DistilBERT, which achieved 99.33% accuracy while being 40% smaller and approximately 60% faster than BERT, with a training time of 1126 seconds compared to 2249 seconds for BERT and 2263 seconds for RoBERTa. This represents nearly a 50% reduction in training time with only a 0.54 percentage point drop in accuracy relative to RoBERTa.
The original version of the BERT achieved the worst performance of the three surveyed models.

TABLE I.
RESULTS

Model	Accuracy	Precision	Recall	F1	Time (s)
RoBERTa	0.9987	0.9966	0.9982	0.9983	2263
DistilBERT	0.9933	0.9831	0.9891	0.9915	1126
BERT	0.9840	0.9604	0.9794	0.9798	2249

VI. Conclusion

The achieved overall results demonstrate the ability of widely available BERT variants with fine-tuning to distinguish between human and AI-generated essays. It is important to point out that the surveyed texts contained significant number of words, providing enough space for the used models to portray their statistical linguistic characteristics, thus enabling easier detection. In the case of shorter texts, such as social media posts, chat messages or even short media articles, the detection would be significantly more difficult, and the accuracy of the classification would drop. Therefore, with future research, we would like to focus on the possibilities of improvement in this field for the short texts. With newly introduced models for generation, we aim to continuously include newer datasets for this task.

Acknowledgement

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, and the Slovak Academy of Sciences under grant no. 1/0685/21 and by the Slovak Research and Development Agency under Contract no. APVV–22–0414.

References

[1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.
[2] Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, *61*(2), 228-239.
[3] A. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, 2023.
[4] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.
[5] Crothers, E. N., Japkowicz, N., & Viktor, H. L. (2023). Machine-generated text: A comprehensive survey of threat models and detection methods. *IEEE Access*, *11*, 70977-71002.
[6] Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Defending against neural fake news. *Advances in neural information processing systems*, *32*.
[7] Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020, July). Automatic detection of generated text is easiest when humans are fooled. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 1808-1822).
[8] Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., ... & Wang, J. (2019). Release strategies and the social impacts of language models. *arXiv preprint arXiv:1908.09203*.
[9] Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023, July). Detectgpt: Zero-shot machine-generated text detection using probability curvature. In *International conference on machine learning* (pp. 24950-24962). PMLR.
[10] Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023, July). A watermark for large language models. In *International Conference on Machine Learning* (pp. 17061-17084). PMLR.
[11] Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). Gltr: Statistical detection and visualization of generated text. *arXiv preprint arXiv:1906.04043*.
[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).
[13] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
[14] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
[15] thedrcat. (2023). *DAIGT V2 Train Dataset* [Dataset]. Kaggle. https://www.kaggle.com/datasets/thedrcat/daigt-v2-train-dataset

Large Reversible Strain in Ni–Fe–Ga Heusler Microwires

[1]*Martin ELIÁŠ (2nd year),*
Supervisor: [2]*Rastislav VARGA*

[1]Dept. of Physics, FEI, Technical University of Košice, Slovak Republic
[1,2]RVmagnetics, a.s., Štúrova 27, 040 01 Košice, Slovak Republic

[1]martin.elias@tuke.sk, [2]varga@rvmagnetics.com

***Abstract*—Glass-coated $Ni_{50}Fe_{27}Ga_{23}$ microwires were fabricated using the Taylor–Ulitovsky method to investigate their potential as magnetic shape memory micro-actuators. A Joule heating screening procedure (controlled current ramp-up and ramp-down with simultaneous resistance monitoring) was implemented to detect phase transitions via characteristic resistance changes, successfully identifying transforming samples. Experimental characterization confirmed a martensite start temperature (M_s) below room temperature and a giant reversible axial strain of 16 % (fractional length change, $\Delta L/L_0$) measured by dynamic mechanical analysis (DMA) during heating under a constant tensile stress of $40\,\mathrm{MPa}$ after cooling to $-100\,^{\circ}\mathrm{C}$. Furthermore, ferromagnetic resonance (FMR) spectra in the martensitic state revealed distinct resonance peaks, possibly corresponding to the easy and hard axes of the martensitic variants.**

***Keywords*—Heusler alloys, Shape Memory Effect, Taylor-Ulitovsky Method, Micro-actuators, Ferromagnetic Resonance.**

I. Introduction and Motivation

The rapid development of micro-robotics, biomedical implants, and smart composite materials has created a critical demand for active components capable of performing mechanical work at the micrometer scale. While glass-coated microwires have traditionally been employed as passive sensing elements, their specific geometry addresses the fundamental limitations of conventional bulk shape memory actuators:

1) **Fast Thermal Response:** The high surface-to-volume ratio facilitates rapid heat dissipation, significantly increasing the actuation frequency bandwidth compared to bulk Shape Memory Alloys (SMAs) which suffer from slow cooling cycles.
2) **Straining in a Single Direction:** The high length-to-diameter ratio ($L/d \gg 1$) confines transformation strain predominantly to the wire axis, making these microwires suitable for linear micro-actuators (e.g., artificial muscles and micropumps).
3) **Embedded Functionality:** With diameters in the range 1–$50\,\mu\mathrm{m}$, comparable to reinforcing fibers, these wires can be embedded into composites without acting as structural defects, enabling "smart materials" with self-sensing and actuation capabilities [1].

Ferromagnetic Shape Memory Alloys (FSMAs) offer a superior actuation mechanism by coupling the structural phase transition with magnetic properties, allowing for non-contact magnetic control [2], [3]. However, the standard Ni–Mn–Ga system suffers from manganese evaporation during the Taylor–Ulitovsky rapid solidification process [4]. This work focuses on the Ni–Fe–Ga system, which offers superior stoichiometric stability, to establish a reproducible fabrication route for magnetic shape-memory microactuators [5], [6].

II. Fabrication and Screening

A. Material Synthesis

Glass-coated crystalline microwires with a nominal composition of $Ni_{50}Fe_{27}Ga_{23}$ were fabricated using the Taylor-Ulitovsky method. The fabricated sample had a metallic nucleus diameter of $d = 14\,\mu\mathrm{m}$ and a total diameter of $D = 48.7\,\mu\mathrm{m}$.

The fabrication process naturally induces substantial internal stresses due to the large difference in thermal expansion coefficients between the metallic core and the glass coating. These stresses manifest primarily as strong axial tension and radial compression on the metallic nucleus. While the glass coating provides environmental protection and structural integrity, its removal is a prerequisite for actuation applications to release the mechanical constraint and allow the core to undergo the macroscopic strain contraction associated with the shape memory effect. However, even after glass removal, residual surface stresses may persist, continuing to influence the final magnetic anisotropy and transformation characteristics [7].

B. Joule Heating Screening

To evaluate functional quality, a Joule heating screening procedure was implemented. The current I was increased and then decreased in controlled steps while continuously monitoring the electrical resistance R. By passing a direct current (DC) through the wire (with glass coating removed), internal heat is generated directly within the core.

As shown in Fig. 1, this method serves as an effective positive indicator for identifying transforming samples. Fig. 1 contains two consecutive heating–cooling cycles measured

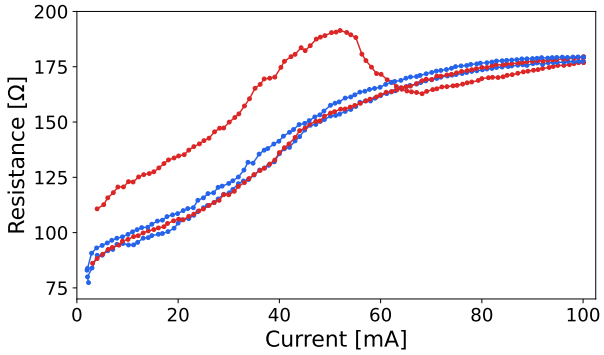


Figure 1. Resistance vs. current ($R(I)$) measured during Joule heating for two consecutive current cycles on the same sample. For each cycle, red denotes the heating branch (current ramp-up) and blue denotes the cooling branch (current ramp-down); thus two red and two blue traces are shown. The non-linear step on heating indicates the martensite-to-austenite transformation.

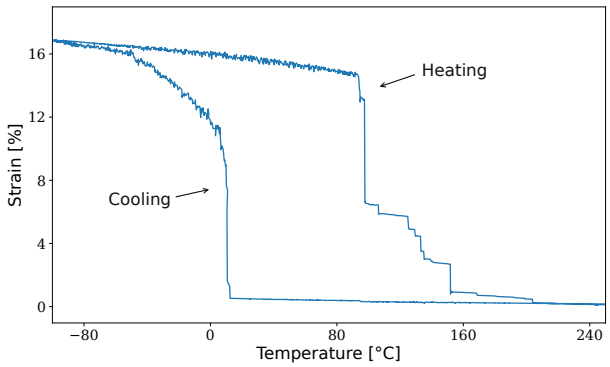


Figure 2. Temperature dependence of strain for a $Ni_{50}Fe_{27}Ga_{23}$ microwire measured via DMA under a constant tensile stress of $40\,\mathrm{MPa}$ during one heating–cooling cycle. The curve shows the thermal hysteresis of the martensitic transformation.

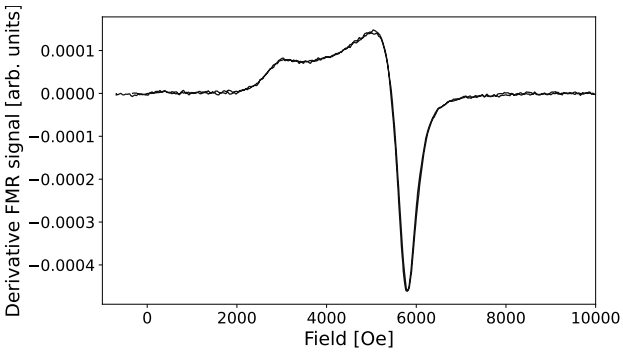


Figure 3. Field-swept FMR spectrum measured at $f = 36\,\mathrm{GHz}$ in the martensitic state, shown as the derivative signal $\mathrm{d}P/\mathrm{d}H$, revealing two resonance peaks.

back-to-back under identical conditions; each cycle consists of a red branch (increasing current) and a blue branch (decreasing current). While a featureless smooth monotonous resistance-current $R(I)$ dependence does not definitively rule out a transformation (which may be suppressed or located outside the accessible temperature range), a characteristic non-linear step provides strong evidence of the phase transition. This characteristic resistance change arises because the crystallographic symmetry of the high-temperature austenite phase (cubic $L2_1$) differs from the low-temperature martensite phase, leading to distinct electron scattering rates.

Notably, the loop does not show a return step to the high-resistance state upon cooling. This is not due to material irreversibility, but rather the experimental boundary conditions: the setup cools only to ambient temperature. Since the martensite start temperature (M_s) of this composition is below room temperature, the reverse transformation is not accessed during the cooling cycle.

III. Experimental Characterization

A. Giant Reversible Strain (DMA)

To quantify the shape memory capability, dynamic mechanical analysis (DMA) was performed in a temperature-controlled chamber. The glass coating was removed from the sample prior to the experiment. The sample was cooled to $-100\,^{\circ}\mathrm{C}$ to ensure a fully martensitic state and then heated under a constant tensile stress of $40\,\mathrm{MPa}$.

Fig. 2 presents the strain–temperature response over one full heating–cooling cycle. During heating, a sharp strain recovery is observed, which coincides with the reverse martensitic transformation between the austenite start and austenite finish temperatures, denoted A_s and A_f, respectively. During subsequent cooling, the strain increases again due to the forward transformation back to martensite and associated variant reorientation. The separation between the heating and cooling branches reflects the thermal hysteresis of the transformation. The analysis yields a giant reversible strain of approximately $16\,\%$. This strain is attributed to a combination of the transformation-related lattice strain and pronounced stress-assisted variant reorientation (detwinning) in the martensitic state, consistent with a strongly oriented crystallographic structure along the wire axis.

B. Magnetic Anisotropy (FMR)

For magnetic actuation, the material must not only transform structurally but also respond strongly to magnetic fields. Complementary ferromagnetic resonance (FMR) measurements at a fixed microwave frequency of $f = 36\,\mathrm{GHz}$ were conducted to probe the magnetic properties of the low-temperature phase.

As shown in Fig. 3, the spectra (shown as the derivative signal $\mathrm{d}P/\mathrm{d}H$) revealed the presence of two distinct resonance peaks in the martensitic state. While definitive assignment requires angle-dependent studies, it can be speculated that these two distinct peaks correspond to the easy and hard axes of magnetization within the martensitic variants. This splitting is consistent with the complex anisotropy landscape typical of multivariant martensite. The presence of multiple resonance conditions suggests that the microwire retains a strong magnetocrystalline anisotropy, a prerequisite for the Magnetic Shape Memory effect (MSM), where a magnetic field reorients the twin variants to produce strain [2], [3].

IV. Summary of Research Activities

Prior to the current focus on Heusler alloys, the research activities focused on the optimization of amorphous bistable microwires for sensor applications.

A key achievement in this area was the identification of an inherent asymmetry between Head-to-Head (HH) and Tail-to-Tail (TT) domain wall (DW) potentials. A mechanical tensile training procedure, defined as a controlled tensile loading–unloading cycle with a maximum load of $100\,\mathrm{MPa}$, was developed, which successfully homogenized the switching field distribution, thereby enhancing sensor precision.

These findings and related activities have been disseminated as follows:

- A manuscript detailing the tensile training effects on DW potentials has been submitted to the *Journal of Magnetism and Magnetic Materials* (Under Review).
- A comprehensive book chapter on contactless sensing has been accepted for publication in *Magnetic Nano- and Microwires* (Elsevier).
- Results were presented at the 18th Czech and Slovak Conference on Magnetism (CSMAG'25).

The current work on Heusler alloys is being conducted during a research stay at the Institute of Physics (FZU) in Prague.

V. Future Plans

Based on the promising actuation potential demonstrated by the Ni–Fe–Ga microwires in this study, future research will prioritize the systematic optimization of their magnetocrystalline anisotropy and phase transformation characteristics. While the primary focus remains on developing functional shape-memory actuators, this work will proceed in parallel with the continued refinement of amorphous sensor materials initiated in previous stages. The research roadmap is divided into two distinct but complementary parts:

A. Optimization of Magnetization Reversal in Amorphous Bistable Microwires

Based on the domain wall (DW) studies described in Section IV, this part focuses on optimizing switching dynamics via thermal treatment, mechanical training, and compositional doping:

- **DW Potential Analysis:** A procedure will be established to extract DW potential shapes from switching-field distributions to quantify the relative contributions of short-range pinning (locally induced anisotropy) versus magnetoelastic pinning.
- **Rare-Earth Doping:** The effect of rare-earth doping (e.g., terbium) on magnetostriction and on the stress dependence of the switching field (H_{sw}) will be quantified, specifically analyzing changes in H_{sw} distributions and stress sensitivity.

B. Development of Functional Shape-Memory Heusler Microwires

The second part focuses on the continued development of the Ni–Fe–Ga system introduced in this work:

- **Material Identification:** Ni–Fe–Ga Heusler samples exhibiting a martensitic transformation will be identified using resistance monitoring during current annealing (Joule heating), followed by the determination of phase constitution, lattice parameters, and composition homogeneity using X-ray diffraction (XRD) and scanning electron microscopy with energy-dispersive spectroscopy (SEM/EDS).
- **Mechanical & Functional Characterization:** Reversible transformation strain and functional fatigue will be determined by DMA under defined preload and thermal cycling. Additionally, magnetic-field-induced strain (MFIS) will be measured under magnetic fields up to $1\,\mathrm{T}$.
- **Magnetic Characterization:** Magnetic anisotropy and easy-axis orientation relative to the wire axis across the transformation will be determined using vibrating sample magnetometry (VSM) measurements ($M(H)$, $M(T)$) together with FMR.
- **Correlation & Processing:** Correlations between fabrication and postprocessing conditions and functional outputs (transformation temperatures, hysteresis, reversible strain, MFIS, saturation magnetization, anisotropy field) will be quantified to define optimal processing conditions.
- **Compositional Tuning:** A major focus will be the precise tuning of the transformation temperatures (M_{s}, A_{f}) and the Curie temperature (T_{c}) via compositional doping. The introduction of dopants such as Cobalt or other transition metals will be investigated to shift the transformation temperatures to desired ranges and enhance magnetic stability. This approach aims to:
  1) **Control T_{t}:** Adjust the structural transformation temperature while maintaining favorable magnetic properties.
  2) **Increase T_{c}:** Ensure T_{c} remains sufficiently high to support ferromagnetic actuation during the entire thermal cycle.

Acknowledgment

This research was supported by the Slovak Research and Development Agency under contract No. APVV-16-0079, and VEGA Grant No. 1/0350/24 from the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic.

References

[1] A. Allue *et al.*, *Composites Part A: Applied Science and Manufacturing*, vol. 120, pp. 12–20, May 2019. DOI: 10.1016/j.compositesa.2019.02.014
[2] J. Dubowik, Y. V. Kudryavtsev, and Y.-P. Lee, *Journal of Magnetism and Magnetic Materials*, vol. 272-276, pp. 1178–1179, May 2004. DOI: 10.1016/j.jmmm.2003.12.1093
[3] L. Kraus *et al.*, 2025. DOI: 10.2139/ssrn.5167726
[4] S. Shevyrtalov *et al.*, *Journal of Alloys and Compounds*, vol. 745, pp. 217–221, May 2018. DOI: 10.1016/j.jallcom.2018.02.096
[5] Y. Li *et al.*, *Scripta Materialia*, vol. 48, no. 9, pp. 1255–1258, May 2003. DOI: 10.1016/S1359-6462(03)00048-4
[6] H. Zheng *et al.*, *Journal of Alloys and Compounds*, vol. 385, no. 1-2, pp. 144–147, Dec. 2004. DOI: 10.1016/j.jallcom.2004.04.124
[7] V. Zhukova *et al.*, *Processes*, vol. 8, no. 8, p. 1006, Aug. 2020. DOI: 10.3390/pr8081006

Emerging Architectures in Modeling and Simulation of Complex Physical Systems using Neural Networks

[1]*Lukáš TOMAŠČÍK (2nd year),*
Supervisor: [2]*Norbert ÁDÁM*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]lukas.tomascik@tuke.sk, [2]norbert.adam@tuke.sk

***Abstract*—Modeling and simulation of complex physical systems is essential in modern science and engineering, yet high-fidelity numerical solvers remain computationally expensive and difficult to deploy in real-time environments. Recent advances in neural networks enable surrogate modeling, solver acceleration, and physics-informed learning strategies that can significantly reduce computational cost. This paper provides an overview of traditional numerical methods and their limitations, a literature review of data-driven, convolutional, graph-based and physics-informed neural approaches, and proposes two central research goals: verifying whether alternative neural architectures can outperform common baselines in the accuracy–efficiency tradeoff for physical simulation, and investigating the integration of pre-trained large language models into the design and configuration of real-time simulation systems.**

***Keywords*—neural networks, machine learning, simulation, physical systems, artificial intelligence**

I. Introduction

Classical simulation of physical systems relies on numerical methods derived from governing equations, typically expressed as systems of ordinary or partial differential equations. Such simulations are indispensable in domains including computational fluid dynamics, structural mechanics, wave propagation and multiphysics modeling. However, high accuracy often implies high computational complexity: dense discretization, fine time stepping and iterative solver loops can result in runtimes incompatible with interactive applications or time-critical decision making.

Neural networks provide an alternative pathway toward efficient modeling by learning to approximate state evolution, closures, or solver operators. Beyond pure approximation, modern research increasingly focuses on hybridization between numerical solvers and learning-based components, as well as physics-constrained learning. A key motivation is enabling simulations in real time, which requires balancing fidelity, stability and inference speed. This work therefore focuses on the analysis of current approaches and on identifying feasible future directions for developing faster and more accurate neural simulation pipelines.

II. Literature Review and Analysis

Established numerical approaches, such as the Finite Element Method (FEM) [1], the Finite Volume Method (FVM) [2], the Lattice Boltzmann Method (LBM) [3], and Smoothed Particle Hydrodynamics (SPH) [4], [5], provide reliable solutions with well-understood stability and convergence properties. They remain dominant in scientific and industrial simulation. However, their computational cost and limited adaptability restrict their use in scenarios where geometry, boundary conditions, or material parameters change frequently. The cost becomes particularly prohibitive when simulations must support optimization loops, parameter sweeps, or online control. Despite their accuracy, classical numerical methods are computationally expensive. High-resolution simulations often involve millions of degrees of freedom and thousands of time steps, resulting in runtimes from minutes to hours. Iterative solvers, such as those for pressure projection or Poisson equations, frequently dominate the cost, while stability constraints further increase computational demands.

The limitations of traditional solvers suggest that future simulation frameworks must explicitly address computational bottlenecks (e.g., pressure projection, iterative Poisson solvers), stability under dynamic domain changes, and deployment constraints under real-time latency requirements.

A direct approach to simulation acceleration is learning surrogates trained on solver-generated data. Early work in graphics demonstrated data-driven fluid simulation using regression forests [6], which illustrates that learned models can replicate complex dynamics under constrained conditions. Similarly, CNN-based solvers and operator approximations were introduced for accelerating Eulerian fluid simulation [7]. Despite excellent speedups, purely data-driven surrogates often struggle with out-of-distribution generalization. In physical simulation, distribution shift is frequent (different geometry, boundary conditions, Reynolds regimes). Therefore, data-driven surrogates are best viewed as components in hybrid pipelines rather than universal replacements for numerical solvers.

Convolutional neural networks (CNNs) naturally encode locality and translational invariance, making them suitable for structured-grid field quantities. They have been applied to pressure prediction, projection steps, and flow field regression [7]. While effective on fixed-resolution grids, CNN methods are constrained by geometry handling and do not directly generalize across unstructured meshes without preprocessing. CNN-based approaches can achieve significant speedups over traditional solvers, with learned components such as pressure projection reducing computation time by up to one or two

orders of magnitude while maintaining acceptable accuracy. However, they often suffer from limited generalization and may degrade on unseen geometries or boundary conditions.

Graph neural networks (GNNs) provide a flexible representation for particle systems, meshes and irregular domains. They support message passing between neighboring entities, mirroring interaction-based physical modeling. Comprehensive reviews have formalized their ability to generalize across graph sizes and topologies [8], [9]. In simulation contexts, GNNs have been used to accelerate Lagrangian fluid simulation [10], offering strong inductive bias for multi-body and multi-phase dynamics. Similarly, differentiable particle simulation networks enable learning of dynamics and physical parameters using gradient-based optimization [11]. Graph-based simulators exhibit strong generalization, particularly in particle systems, where models trained on small setups can scale to larger ones with comparable accuracy. They also tend to show improved stability and lower long-term error, though at higher per-step computational cost.

Physics-informed neural networks (PINNs) incorporate governing equations into the learning objective via automatic differentiation [12], [13]. PINNs reduce dependence on labeled simulation data and allow inverse problems (parameter estimation and hidden state reconstruction). A broader framework for physics-informed machine learning is presented in [14].PINNs excel at solving inverse problems and low-dimensional PDEs with high accuracy, often without labeled data. However, they require long training times and careful hyperparameter tuning to avoid stagnation or loss imbalance. Challenges include instability, stiffness, and sensitivity to initialization. Robust training strategies and hybrid integration with physical constraints are active research directions.

Real-time deployment often requires reduction in model size and inference cost. Pruning [15] eliminates redundant weights or layers, while knowledge distillation [16] transfers the behavior of a large teacher network into a compact student. These techniques enable deployment under limited memory and compute budgets. Reinforcement learning (RL) introduces a complementary approach: learning control and adaptation policies interacting with simulators [17], [18]. In hybrid frameworks, RL can be used to tune solver parameters, select fidelity levels, or guide adaptive sampling. This direction becomes particularly relevant when simulation is part of a closed-loop system.

A promising direction is integrating pretrained Large Language models (LLMs) to guide simulation design. Unlike traditional surrogates, they act as meta-models, assisting with workflow generation, adaptive parameter tuning, and solver configuration.This aligns with the second main research goal: investigating whether LLM integration can increase speed and accuracy of real-time simulation workflows, particularly when combined with PINN-based approaches.

III. Simulation Acceleration Opportunities and Research Perspective

The literature review confirms that neural simulation is already an established field, with strong results in surrogate modeling, solver acceleration, and physics-informed learning. Nevertheless, the majority of existing implementations remain surprisingly conservative in their architectural choices: CNN-based models dominate grid-based surrogate learning, GNNs are increasingly standard for particle methods, and classical multilayer perceptrons remain the default for physics-informed neural networks. This creates an opportunity for research that is not merely incremental, but structural, namely by revisiting alternative neural architectures and by introducing new hybrid workflows in which pretrained large language models contribute to physics-informed learning.

This section outlines two research directions aligned with the primary goals: investigating the potential of alternative network designs, including architectures that are currently underutilized in physical simulation, and examining how LLMs can be integrated into PINN pipelines to improve training stability and real-time usability.

A large portion of modern neural simulation research can be interpreted as a search for architectures with strong inductive bias: locality for CNNs, relational interactions for GNNs, or differential constraints for PINNs. However, there is still a gap between expressive power and efficient expressiveness. Many physical systems contain structured relationships that do not necessarily require arbitrarily deep networks but may benefit from architectures that encode compositional structure more explicitly.

One example of this idea can be found in networks motivated by the Kolmogorov–Arnold representation theorem. Modern implementations inspired by this theorem, often referred to as Kolmogorov–Arnold Networks (KANs), represent multivariate mappings through compositions of learnable univariate functions rather than using only affine transformations followed by pointwise nonlinearities [19]. While such models are not yet common in simulation practice, they suggest a valuable hypothesis for real-time modeling: certain physical dependencies may be approximated more compactly by structured nonlinear compositions than by dense layers, potentially improving accuracy while reducing parameter count and inference cost.

Beyond KAN-inspired models, the broader idea of revisiting less mainstream architectures is relevant precisely because physical simulation problems rarely resemble generic machine learning datasets. The approximation targets tend to be smooth, constrained by conservation laws, and strongly governed by local interactions. In this context, models that explicitly reflect smoothness, locality, and structure may achieve better accuracy–efficiency tradeoffs than standard baselines. From the author's perspective, this makes the systematic benchmarking of alternative architectures a meaningful and feasible contribution, particularly when combined with constraints of real-time deployment.

A second promising direction is the integration of pretrained LLMs into simulation pipelines, particularly those based on physics-informed neural networks. LLMs should not be interpreted as replacements for numerical computation; rather, their strength lies in their ability to operate as reasoning and configuration systems. They can assist in specifying problems, adapting workflows, and controlling training procedures, which is highly relevant because PINNs are often limited not by the representational capacity of networks but by the difficulty of optimization and formulation.

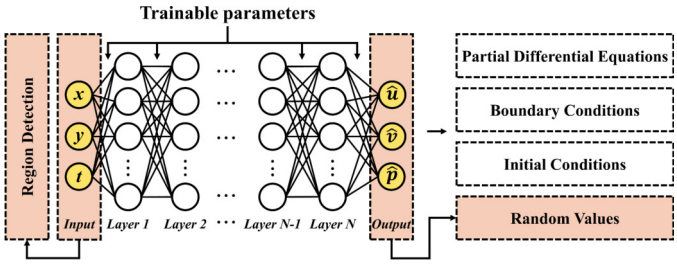


Fig. 1. Random reinitalization strategy of PINN [20]. This illustrates the sensitivity of PINNs to various intial factors

PINN training is well known [20] to be sensitive to the choice of collocation points, the weighting between PDE residual terms and boundary or initial constraints, and the stability of optimization across different scales of the solution as illustrated in Fig. 1. In practice, successful PINN deployment typically requires extensive manual tuning and repeated experimentation. This tuning process becomes even more problematic in real-time or near-real-time scenarios, where problem conditions can change and retraining must be efficient and reliable. LLM-augmented PINNs can enhance adaptability by interpreting PDEs and boundary conditions, suggesting training configurations, adjusting loss weights or sampling, and identifying symmetries or constraints to stabilize training. The LLM in the proposed framework does not approximate the physical system directly, but supports the training of physics-informed neural networks. Given a PDE-based problem formulation, an initial PINN is constructed and trained while monitoring loss components and residual errors. This information is provided to the LLM, which suggests adjustments such as loss reweighting, sampling strategies, or normalization changes. The training is iteratively updated based on these suggestions until convergence, allowing the LLM to act as an adaptive component that improves consistency and reduces manual tuning.

We hypothesize that the integration of pretrained LLMs into PINN-based simulation pipelines can improve practical usability and potentially increase solution quality in real-time contexts. The expected advantage lies in reducing the amount of manual tuning required for stable PINN convergence and enabling more autonomous adaptation of training pipelines.

The research will be tested on benchmark PDE problems, such as incompressible flow and diffusion–advection, enabling direct comparison of numerical solvers and neural methods. Performance metrics include accuracy (e.g., MSE, relative L^2 norm), efficiency, and stability. Different neural architectures will be compared to MLP and CNN baselines, and LLM-augmented PINNs will be evaluated for convergence, robustness, and stability. Currently, the focus is on analysis and problem formulation and experiments are planned for future validation.

IV. Conclusion

Recent results demonstrate strong potential for accelerated simulation, while the analysis indicates that future progress depends not only on scaling networks or increasing training data, but also on developing more structured simulation pipelines and making neural solvers more robust and deployable.

Based on the analysis and initial progress, two feasible and impactful research directions are identified. First, systematically exploring and benchmarking alternative neural architectures, including models inspired by Kolmogorov–Arnold representations, to improve the accuracy–efficiency tradeoff in simulations. Second, developing LLM-augmented PINN workflows, where pretrained language models assist in configuring, stabilizing, and adapting physics-informed training to support real-time simulations.

The planned work will therefore focus on evaluating alternative architectures for neural simulation and on designing hybrid frameworks that combine PINNs with LLM-driven configuration mechanisms, aiming to improve accuracy, stability, and computational efficiency of simulation pipelines under real-time constraints. The identified challenges and proposed directions are derived from a critical analysis of the literature and observed limitations reported across multiple studies.

References

[1] J. N. Reddy and D. K. Gartling, *The finite element method in heat transfer and fluid dynamics*. CRC press, 2010.
[2] R. Eymard, T. Gallouët, and R. Herbin, "Finite volume methods," *Handbook of numerical analysis*, vol. 7, pp. 713–1018, 2000.
[3] Y. B. Bao and J. Meskas, "Lattice boltzmann method for fluid simulations," *Department of Mathematics, Courant Institute of Mathematical Sciences, New York University*, vol. 44, 2011.
[4] R. A. Gingold and J. J. Monaghan, "Smoothed particle hydrodynamics: theory and application to non-spherical stars," *Monthly notices of the royal astronomical society*, vol. 181, no. 3, pp. 375–389, 1977.
[5] L. B. Lucy, "A numerical approach to the testing of the fission hypothesis," *Astronomical Journal, vol. 82, Dec. 1977, p. 1013-1024.*, vol. 82, pp. 1013–1024, 1977.
[6] L. Ladickỳ, S. Jeong, B. Solenthaler, M. Pollefeys, and M. Gross, "Data-driven fluid simulations using regression forests," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 1–9, 2015.
[7] J. Tompson, K. Schlachter, P. Sprechmann, and K. Perlin, "Accelerating eulerian fluid simulation with convolutional networks," in *International conference on machine learning*. PMLR, 2017, pp. 3424–3433.
[8] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI open*, vol. 1, pp. 57–81, 2020.
[9] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay, "Graph neural networks," *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 17, 2024.
[10] Z. Li and A. B. Farimani, "Graph neural network-accelerated lagrangian fluid simulation," *Computers & Graphics*, vol. 103, pp. 201–211, 2022.
[11] C. Schenck and D. Fox, "Spnets: Differentiable fluid dynamics for deep neural networks," in *Conference on Robot Learning*. PMLR, 2018, pp. 317–335.
[12] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10561*, 2017.
[13] M. Raissi, "Deep hidden physics models: Deep learning of nonlinear partial differential equations," *Journal of Machine Learning Research*, vol. 19, no. 25, pp. 1–24, 2018.
[14] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
[15] D. Blalock, J. J. Gonzalez Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?" *Proceedings of machine learning and systems*, vol. 2, pp. 129–146, 2020.
[16] G. Hinton, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.
[17] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.
[18] J. Rabault, F. Ren, W. Zhang, H. Tang, and H. Xu, "Deep reinforcement learning in fluid mechanics: A promising method for both active flow control and shape optimization," *Journal of Hydrodynamics*, vol. 32, no. 2, pp. 234–246, 2020.
[19] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.
[20] J. Lee, S. Shin, T. Kim, B. Park, H. Choi, A. Lee, M. Choi, and S. Lee, "Physics informed neural networks for fluid flow analysis with repetitive parameter initialization," *Scientific Reports*, vol. 15, no. 1, p. 16740, 2025.

Hybrid FSO/RF Link Management Using RSSI Prediction and Reinforcement Learning

[1]*Zuzana LIŠČINSKÁ (2nd year),*
Supervisor: L'uboš OVSENÍK

[1,2]Dept. of Computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]zuzana.liscinska@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*—This paper presents an integrated framework for hybrid FSO/RF link management based on short-horizon RSSI (Recieved Signal Strength) prediction and reinforcement learning. Real minute-resolution measurements from a meteorological station at the Technical University of Košice was used to train and evaluate sequence models LSTM (Long Short Term Memory) and Transformer. The predicted RSSI values were subsequently incorporated into a hard switching mechanism between FSO and RF links, implemented using a DQL (Deep Q-Learning) agent and compared with a threshold-based baseline ($\mathbf{RSSI_{pred} < -30}$ dBm). Experimental results indicate that sequence-based prediction models accurately capture baseline RSSI behavior and gradual variations, while reinforcement learning enables more stable and proactive switching decisions, particularly during short-term signal degradations.**

***Keywords*—DQL, FSO, Hybrid FSO/RF, LSTM, RF, RSSI, Transformer**

I. Introduction

Hybrid FSO/RF links are widely considered as a practical solution for maintaining high-capacity connectivity under rapidly changing atmospheric conditions. While the FSO (Free Space Optics) channel offers very high data rates, its performance can degrade sharply during fog, low visibility, precipitation, or strong turbulence, which leads to significant attenuation and unstable received optical power. To mitigate these outages, a parallel RF (Radio Frequency) link is typically used as a backup channel, improving link availability and service continuity [1].

A key practical challenge is not only the existence of a backup channel, but the timing and correctness of switching decisions. In hard switching, the system selects either FSO or RF at a given time. Late switching may cause packet loss or temporary service interruption, whereas premature switching can unnecessarily reduce throughput and spectral efficiency. For this reason, reliable link management should not rely solely on the current RSSI value, but should incorporate short-horizon prediction of link quality to anticipate imminent degradations [2], [3].

This paper presents the work conducted during the last academic year on an integrated framework combining minute-ahead RSSI prediction with hard switching decisions in hybrid FSO/RF systems. Meteorological inputs are used to forecast RSSI via sequence models LSTM and Transformer, and the predicted values are subsequently applied to channel selection using reinforcement learning DQL. The main objective is to evaluate the impact of prediction accuracy on switching performance, particularly during rapid RSSI drops, and to establish a reproducible framework for future system-level optimization.

II. Baseline Framework And Research Workflow

To ensure that switching decisions can be evaluated in a realistic and reproducible way, the work conducted during the last academic year focused on building a complete end-to-end pipeline. This section summarizes the established baseline: the data strategy, the implemented RSSI prediction models, and the way predictions are translated into hard switching decisions.

A. Real Measurements and Synthetic Data Generation

Real minute-resolution measurements were collected at TUKE to capture real operational behavior under varying atmospheric conditions. The dataset combines RSSI values with meteorological variables, including temperature, humidity, pressure, particle concentration, visibility, and wind speed. These measurements provide a realistic basis for evaluating link performance and switching behavior under naturally occurring environmental variations.

B. RSSI Prediction Using Sequence Models

The prediction task was formulated as a short-horizon time-series forecasting problem with a one-minute horizon, reflecting the operational need to anticipate imminent degradations before they translate into outages or unnecessary switching [4]. The input to the predictor consists of multivariate sequences built from meteorological and environmental variables together with recent RSSI history, so that the model can learn both slow trends (e.g., daily or seasonal patterns) and fast local fluctuations. In addition to the one-step prediction, the same setup supports iterative multi-step use for longer scenario simulations.

Two model branches were implemented and compared to understand how different sequence-learning mechanisms behave under realistic and critical conditions. The first branch employs an LSTM architecture, selected for its ability to capture temporal dependencies through gated recurrent dynamics. LSTM is suitable for RSSI prediction because it can capture time dependencies and long-term trends in sequential data, which is important when modeling gradual signal changes affected by meteorological conditions. It works on the principle of a recurrent neural network with memory cells and control gates (input, forget, output gate) that selectively retain

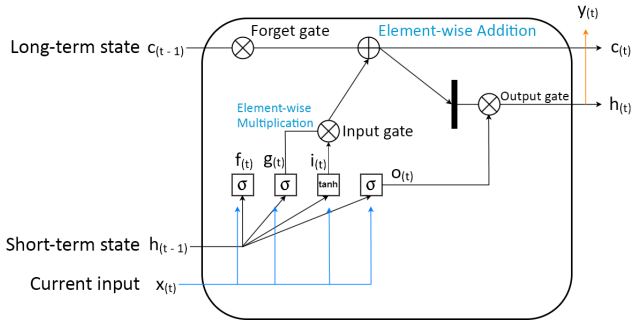


Fig. 1. LSTM architecture cell

or suppress information from previous time steps [5]. The Fig. 1 shows the architecture of the LSTM model cell.

The second branch investigates a Transformer-based approach, motivated by its attention mechanism and its capacity to model complex dependencies within the input window. The goal is to improve stability and responsiveness when the signal dynamics change rapidly, especially during short critical intervals linked to visibility degradation [6]. Besides learning the absolute RSSI level, experiments also consider predicting the incremental change (e.g., $\Delta RSSI$) to emphasize transition behavior and reduce systematic smoothing of sharp drops. This formulation is particularly relevant for switching, where correctly anticipating the onset and depth of a decline can be more important than minimizing average error in stable periods. The Fig. 2 shows the architecture of the Transformer model.

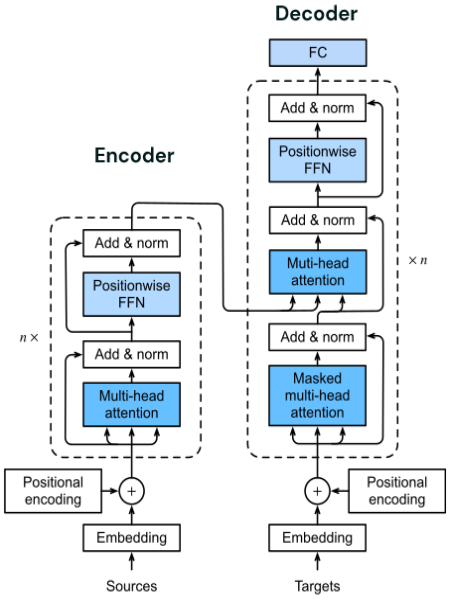


Fig. 2. Transformer architecture

C. From Prediction to Hard Switching Decisions

To connect forecasting with link management, the predicted RSSI is integrated into a decision-making layer responsible for *hard switching* between the FSO and RF channels, i.e., selecting exactly one active link at each time step. In this workflow, the predictor provides a short-horizon estimate of the upcoming channel quality, which allows the controller to act proactively instead of reacting only after the RSSI has already dropped [7]. This is especially relevant in short critical intervals, where fast degradations may cause outages if the switching decision is delayed, while overly conservative switching can unnecessarily reduce throughput by falling back to RF too early.

A reinforcement learning agent based on DQL was adopted to learn the channel-selection policy through repeated interaction with the simulated environment. At each decision step, the agent observes a state vector composed of recent RSSI history, selected meteorological variables, and the predicted RSSI, and chooses a discrete action corresponding to activating either the FSO or the RF link. The reward design reflects both communication performance and operational stability. It encourages maintaining the channel with higher expected quality and availability, while penalizing excessive switching to suppress the *ping-pong* effect. This formulation aims to achieve a practical trade-off between reliability and efficiency [8].

III. Research Outcomes and Key Milestones

This section summarizes the main outcomes achieved during the last academic year. The reported results are organized to reflect the complete workflow from data preparation to decision-making.

A. Evaluation of prediction results using LSTM model

In this subsection, we evaluate the forward *minute* prediction of the RSSI parameter using the LSTM model. Graph on Fig. 3 shows a comparison of the actual RSSI (red curve) and the predicted RSSI (blue curve) during the day 08.12.2025. It is clear from the graph that the model most of the time faithfully copies the basic signal level and captures its gradual changes in stable sections, where the curves almost overlap. Larger deviations appear during short-term sharp RSSI drops, which the LSTM often *smooths* (predicts a milder drop) and in some cases reacts with a slight delay.

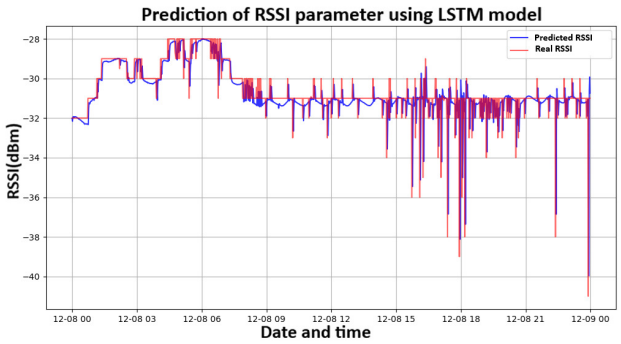


Fig. 3. Prediction of RSSI parameter using LSTM model

TABLE I
Excerpt of a critical section with predicted RSSI and switching between FSO and RF transmission links

Date and Time	$RSSI_{pred}$	RSSI	Decision
12/8/2025 08:00	-31.1267	-31	0
12/8/2025 08:01	-30.3142	-31	1
12/8/2025 08:02	-30.7678	-31	0
12/8/2025 08:03	-31.1255	-30	0
12/8/2025 08:04	-31.1110	-30	0
12/8/2025 08:05	-30.3340	-30	1

Table I presents an example of RSSI parameter prediction on 08.12.2025 in the time interval from 08:00-08:05. The model

was trained on real data obtained from a meteorological station located at the Technical University campus, with the data split into training and validation sets in an 80:20 ratio. A value of "1" in the "Decision" column indicates signal transmission via the FSO link, while a value of "0" denotes transmission via the RF link when the predicted RSSI is below -30 dBm.

B. Evaluation of prediction results using Transformer model

The graphical comparison of the predicted and actual RSSI values shown in Fig. 4 indicates that the Transformer model accurately captures the baseline signal level and follows its gradual variations throughout most of the day. During the morning period, when the RSSI increases from approximately -32 dBm to around -28 dBm, the Transformer correctly reflects both the overall trend and the step-like nature of the changes. The most challenging segments are short-term sharp drops in RSSI, particularly in the afternoon and evening hours, where the model generally detects the decreases but does not always predict their exact depth, occasionally overestimating or underestimating their magnitude.

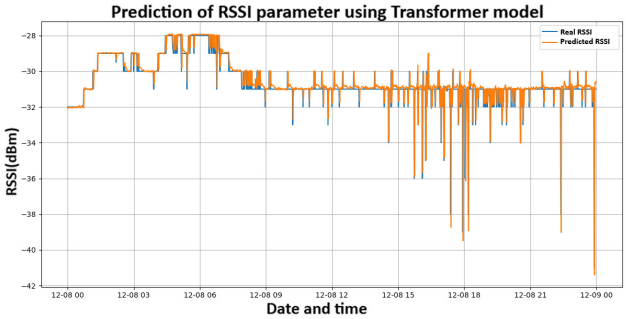


Fig. 4. Prediction of RSSI parameter using Transformer model

TABLE II
EXCERPT OF A CRITICAL SECTION WITH PREDICTED RSSI AND SWITCHING BETWEEN FSO AND RF TRANSMISSION LINKS

Date and Time	$RSSI_{pred}$	RSSI	Decision
12/8/2025 08:00	-29.9873	-31	1
12/8/2025 08:01	-30.7918	-31	0
12/8/2025 08:02	-30.7151	-31	0
12/8/2025 08:03	-30.7358	-30	0
12/8/2025 08:04	-29.9654	-30	1
12/8/2025 08:05	-29.9753	-30	1

Table II presents a brief example of RSSI parameter prediction for 08.12.2025 in the time interval from 08:00 to 08:05. The prediction model was trained on real measured data obtained from a meteorological station located on the Technical University campus, with the data divided into training and validation sets in an 80:20 ratio. A value of "1" in the "Decision" column indicates signal transmission via the FSO link, while a value of "0" denotes transmission via the RF link when the predicted RSSI value falls below -30 dBm.

C. Research Achievements and Project Recognition

The research conducted within the dissertation topic has resulted in several peer-reviewed publications indexed in SCOPUS and Web of Science. In 2024, the paper entitled "Deep Data Analysis Methods Applied for Hard Switching in Hybrid FSO/RF Systems" was published in the proceedings of ICETA 2024. In 2025, the contribution "The Importance of Correlation Coefficients in Optimizing Input Data for Hybrid FSO/RF Systems" was presented at the Radioelektronika 2025 conference and indexed in SCOPUS and WoS databases.

Further research outputs include an accepted journal paper in *Acta Electronica et Informatica* (in press) and accepted and presented contributions at ICETA 2025 and CINTI 2025 international conferences.

The research activities are additionally supported by the successful acquisition of the Early Stage Grant (Project No. 09I03-03-V05-00015), where the author serves as the principal investigator. The project focuses on the application of artificial intelligence methods for intelligent hard switching between FSO and RF links under varying atmospheric conditions, aiming to improve transmission reliability and system performance.

IV. CONCLUSION

The proposed framework demonstrates that combining minute-ahead RSSI prediction with reinforcement learning provides an effective approach to proactive link management in hybrid FSO/RF systems. Both LSTM and Transformer models successfully capture dominant signal trends, with deviations mainly occurring during short-term sharp drops. Integrating predicted RSSI into a DQL-based decision layer improves switching stability compared to a fixed threshold strategy, especially under rapidly changing atmospheric conditions. In the next stage of research, it is planned to implement a PPO (Proximal Policy Optimization) agent for adaptive control of hard handover between the FSO and RF links.

ACKNOWLEDGMENT

This work was supported by following research grant: Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00015.

REFERENCES

[1] R. Rani, A. Mandpura, K. Chauhan, D. Yadav, and M. Lakshmanan, "Performance of adaptive combining-based hybrid fso/rf communication system," *Photonic Network Communications*, vol. 51, 2025.

[2] M. Mansour Abadi, Z. Ghassemlooy, S. Zvanovec, M. Bhatnagar, and Y. Wu, "Hard switching in hybrid FSO/RF link: Investigating data rate and link availability," *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017.

[3] K. Sharma, S. Kaur, and H. Singh, "Channel modelling and performance analysis of switching based hybrid fso/rf communication system," *Journal of Optics*, 2025.

[4] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, 2014.

[5] Y. Lin, W. Li, J. Sun, and Q. Wu, "Improving wireless devices identification using gray relationship classifier to enhance wireless network security," *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2018.

[6] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Beyond efficient transformer for long sequence time-series forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021.

[7] S. Henna, A. A. Minhas, M. S. Khan, and M. S. Iqbal, "Ensemble consensus representation deep reinforcement learning for hybrid FSO/RF communication systems," *Optics Communications*, vol. 530, 2023.

[8] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, and L.-C. Wang, "Deep reinforcement learning for mobile 5G and beyond: Fundamentals, applications, and challenges," *IEEE Vehicular Technology Magazine*, vol. 14, 2019.

Current State and Innovative Concepts in Technological Support for Diagnosis and Therapy of Speech and Hearing Disorders

[1]*Juraj KOVÁČ (1st year)*
Supervisor: [2]Stanislav ONDÁŠ

[1,2]Dept. of Computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]juraj.kovac.3@tuke.sk, [2]stanislav.ondas@tuke.sk

***Abstract*—This paper reviews state-of-the-art technologies for speech and hearing disorders, focusing on AI diagnostics, gamified therapy, and tele-audiology. The analysis reveals a critical lack of robust models for inflective languages like Slovak and limited adaptivity in current tools. To address these gaps, the paper proposes a modular, human-centered AI ecosystem that integrates speech and hearing rehabilitation through dynamic, personalized intervention.**

***Keywords*—Adaptive Systems, Artificial Intelligence, Gamification, Hearing Rehabilitation, Speech Therapy, Telemedicine.**

I. Introduction

Communication plays a pivotal role in social interaction and cognitive development. However, speech and hearing disorders represent a significant barrier for a substantial portion of the population, affecting quality of life, education, and social inclusion. Conventional methods of speech-language pathology and audiology face several challenges. Firstly, the therapeutic process is resource-intensive and often limited by the shortage of qualified therapists. Secondly, traditional diagnostics can suffer from subjectivity, relying heavily on the clinician's auditory perception and experience. Finally, specifically in pediatric care, maintaining long-term patient adherence to repetitive therapeutic exercises is difficult due to a lack of motivation and engagement [1].

In response to these challenges, recent years have seen a surge in integrating modern technologies into clinical practice. Concepts such as serious games (gamification), artificial intelligence (AI), and telemedicine are emerging as powerful tools to support both diagnostics and therapy. While gamification addresses the engagement gap by providing immediate feedback and reward systems, machine learning models offer the potential for objective, automated assessment of speech pathologies. Despite the growing number of studies in these isolated areas, there is a need for a comprehensive review that bridges the gap between engineering solutions and clinical applicability [1] [2].

To systematically map the current landscape, this review was guided by the following primary research question: How are current artificial intelligence, gamification, and telemedicine technologies integrated into the diagnosis and therapy of speech and hearing disorders, and what are their specific limitations regarding adaptivity and morphologically rich languages? The literature search focused on recent peer-reviewed publications (predominantly from 2022 to 2026), targeting keywords such as AI-based speech diagnostics, serious games in audiology, and tele-rehabilitation.

This paper provides a critical review of the current state of knowledge in the technological support for speech and hearing disorders. It analyses existing approaches in automated diagnostics and gamified therapy, identifies limitations in current methodologies, particularly regarding language-specific dependencies and outlines future directions for creating more adaptive, personalized, and accessible healthcare solutions.

II. Review of Current Technological Approaches in Speech and Hearing Care

The integration of technology into speech-language pathology (SLP) and audiology has created a complex research landscape. Current academic literature indicates a split in technological applications: those aimed at diagnostic support, striving to replace or augment subjective clinician judgments with objective metrics, and those focused on therapeutic intervention, addressing patient's motivation and accessibility. The following sections critically analyse the state-of-the-art in these domains, categorized by the core technologies employed ranging from artificial intelligence and machine learning signal processing to gamification and tele-rehabilitation platforms.

A. AI & Machine Learning

Recent advancements in deep learning have revolutionized the analysis of pathological speech. State-of-the-art models, particularly Convolutional Neural Networks (CNNs) and transformer-based architectures like Wav2Vec 2.0, have demonstrated superior performance in detecting acoustic biomarkers for conditions such as dysarthria, stuttering, and speech sound disorders. Unlike traditional methods dependent on subjective auditory-perceptual analysis, these algorithms can identify subtle non-linear patterns in speech signals with high accuracy, often exceeding 90% accuracy for English datasets in controlled, noise-free clinical environments [3] [4].

However, the clinical adoption of these "black-box" models remains limited due to interpretability challenges and ethical concerns regarding algorithmic bias and data privacy [5]. Pure automation cannot replicate the empathy and contextual

understanding of a human therapist. Instead, the current research paradigm is shifting towards Human-Centered AI (HCAI). This approach emphasizes systems that do not aim to replace the clinician but rather augment their capabilities providing automated pre-screening and phonological error detection while maintaining a high level of human control over the final diagnosis [6]. This "human-in-the-loop" strategy addresses the trust gap by ensuring that AI serves as a transparent decision-support tool rather than an autonomous evaluator, a crucial requirement for integration into sensitive pediatric healthcare workflows [7].

B. Gamification and Immersive Serious Games

While automated diagnostics focus on the clinician's needs, gamification addresses the patient's primary challenge: adherence to repetitive therapy. Traditional speech and auditory training often suffer from low engagement, particularly in pediatric populations. To mitigate this, researchers have increasingly adopted "Serious Games" which are digital interventions where the primary goal is therapeutic rather than entertainment [8].

Initial implementations focused on 2D mobile and web applications providing immediate rewards and visual feedback. However, a critical analysis of commercial applications reveals a significant limitation: they often lack configurability for specific patient needs. As proposed in platform concepts like LogoHra [9], the effectiveness of

Fig. 1. Showcase of application LogoHra.

gamified therapy relies heavily on the "therapist-in-the-loop" architecture. This allows clinicians to customize lexical content and difficulty levels, ensuring that the game complements rather than replaces clinical sessions. Furthermore, for hearing-impaired children, gamification must go beyond simple audio playback. Recent studies evaluating tools for the TEKOS II test demonstrate that incorporating rich storytelling and visual-only stimuli is essential for accurately assessing comprehension in children with cochlear implants or hearing aids, where auditory discrimination is compromised [10].

The current state-of-the-art are shifting from screen-based interaction to immersive Virtual Reality (VR). Unlike tablets, VR eliminates external distractions and allows for 3D visualization of articulatory movements. Systems like EchoVR utilize avatars to demonstrate lip rotations and tongue placement, providing a "digital mirror" for the patient. This is particularly effective for dyslalia and motor-speech disorders, as children can mimic an avatar (e.g., "Dex") in a controlled, anxiety-free environment. Similarly, frameworks integrating Machine Learning with VR, such as LingoLand are beginning to offer real-time pronunciation feedback, bridging the gap between static gaming and dynamic correction [11] [12].

Despite these advancements, a gap remains between engagement and efficacy. Most current games use static difficulty settings. The future of therapeutic gaming lies in Dynamic Difficulty Adjustment (DDA) systems that use real-time performance data to automatically adjust the complexity of the game (e.g., background noise levels for hearing training or word complexity for speech), a concept that remains under-explored in current Slovak-language solutions.

C. Virtual and Augmented Reality (VR/AR)

The transition from 2D screen-based applications to immersive environments represents a significant leap in therapeutic technology. Virtual Reality (VR) and Augmented Reality (AR) offer a safe, controlled setting where patients can practice communication skills without the fear of real-world judgment [13].

For speech disorders with a psychological component, such as stuttering, VR serves as a powerful exposure therapy tool. Recent studies from 2025 highlight the efficacy of Virtual Reality Exposure Therapy (VRET), where patients navigate simulated social scenarios such as ordering food in a noisy café or speaking publicly while a therapist monitors their anxiety levels and speech fluency in real-time [14].

Unlike traditional therapy, which relies on role-playing, VR provides realistic environmental distractors (background noise, visual crowds), making the transfer of skills to daily life more effective. As demonstrated in the EchoVR system, the use of animated avatars (e.g., "Dex") can further clarify articulatory mechanics for children, visualizing hidden tongue movements that are impossible to show in a mirror [11].

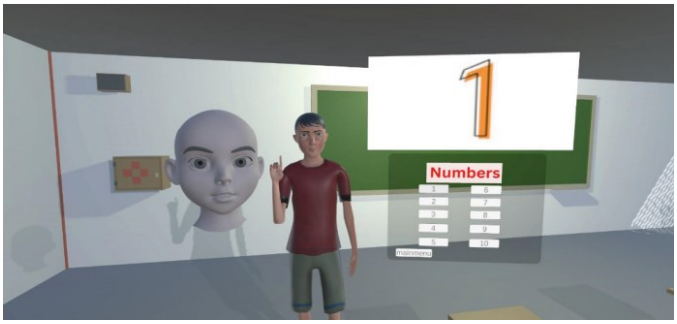


Fig. 2. Showcase of EchoVR.

While VR isolates the user, AR enhances the real world, making it particularly suitable for hearing rehabilitation. A promising innovation is the use of consumer-grade AR smart glasses (e.g., XREAL, Epson) to provide real-time speech-to-text transcription directly in the user's field of view. This technology acts as a "visual hearing aid," allowing individuals with severe hearing loss to follow conversations by reading captions floating above the speaker's head. Furthermore, for children with cochlear implants, AR applications can gamify the auditory training process by overlaying virtual rewards onto physical objects in the room, encouraging active listening and sound localization in their natural environment [15] [16].

Despite the potential widely accessible implementation faces hurdles. High hardware costs and the "cybersickness" effect remain barriers for widespread home use. Moreover, a

systematic review notes that while pilot studies show high engagement, there is a lack of long-term longitudinal data confirming that skills acquired in VR persist after the novelty wears off [13].

D. Technologies for Diagnosis and Rehabilitation of Hearing Disorders

While speech disorders often manifest externally, hearing impairments represent a sensory deficit requiring different technological interventions. The focus here shifts from output correction (speech) to input enhancement and auditory training.

The traditional diagnosis of hearing loss, relying on Pure Tone Audiometry (PTA) in sound-treated booths, is a bottleneck in global healthcare. Recent trends favor decentralized diagnostics via smartphone applications. Validation studies from 2024 indicate that calibrated mobile apps (e.g., using bundled consumer headphones) can achieve screening accuracy of over 95% compared to clinical audiometers. These "pocket audiometers" utilize active noise monitoring algorithms to reject test results if ambient noise levels exceed a threshold, ensuring reliability even in home environments. For a dissertation focusing on accessibility, integrating a pre-screening hearing module into a speech therapy platform is a logical step, as undetected hearing loss often underlies speech delays [15].

Hardware limitations are being overcome by edge-AI. The latest generation of hearing aids employs on-chip Deep Neural Networks (DNNs) to separate speech from noise in real-time, replacing standard directional microphones. For Cochlear Implant (CI) users, the challenge is different: electrical stimulation provides a degraded spectral signal. Here, technology focuses on "Music-induced Rehabilitation." Since CI users struggle with pitch perception, new adaptive algorithms amplify rhythmic cues to aid music appreciation, which has been shown to cross-transfer skills to prosodic speech perception [16] [17].

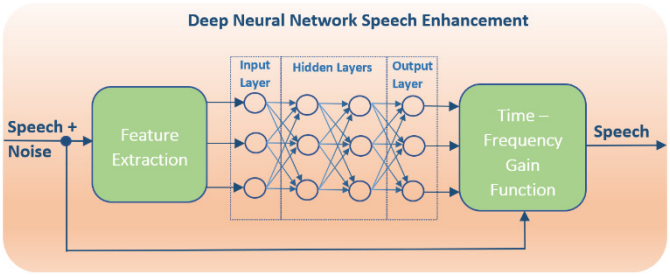


Fig. 3. Deep Neural Network Speech Enhancement schema.

Post-implantation rehabilitation is critical but tedious. Conventional auditory training (listening to beep sequences) lacks engagement for children. The solution lies in "Serious Games" that visualize sound. For example, pitch training can be gamified by controlling a character's vertical movement through voice pitch (high pitch = jump up). Recent trials with the *World of Sounds* game demonstrated that CI recipients using gamified training showed significantly faster improvements in receptive language skills compared to control groups using traditional methods. This directly validates the concept of combining speech and hearing exercises into a single, gamified ecosystem [18].

Beyond diagnostics, the management of hearing devices is shifting towards Remote Fitting. Traditionally, users of Cochlear Implants (CI) or hearing aids had to visit a clinic physically for device tuning (mapping). New protocols introduced in 2024 allow audiologists to adjust frequency gain and compression parameters asynchronously via cloud-based platforms. The patient uses a mobile app to log specific listening challenges (e.g., "speech in noise is unclear"), and the system uploads data-logs to the clinician. This "store-and-forward" tele-audiology model not only reduces hospital visits but allows for fine-tuning based on real-world acoustic environments rather than artificial clinic conditions [18].

A critical innovation for patients with severe hearing loss is the integration of visual cues into auditory processing. It is well documented that visual speech perception (lip-reading) significantly improves speech intelligibility in noisy environments. Emerging assistive technologies now utilize Audio-Visual Speech Recognition (AVSR) systems running on smartphones or AR glasses. These systems use the device's camera to track the speaker's lip movements and use this visual data to filter the audio signal more aggressively than audio-only algorithms could. For a dissertation focusing on both speech and hearing, this represents a convergence point: the same computer vision models used to detect dyslalia (in speech therapy) can be repurposed to enhance audio signals for the hearing impaired [19].

E. Telemedicine and Remote Monitoring

The integration of telemedicine into speech-language pathology and audiology has shifted from being a supplementary service to a core modality of care, particularly following the COVID-19 pandemic. Current literature distinguishes between synchronous telepractice (real-time video conferencing) and asynchronous remote monitoring (store-and-forward data), both of which are essential for a holistic therapeutic ecosystem.

Extensive systematic reviews have validated that telepractice is non-inferior to in-person therapy for a wide range of pediatric speech disorders. A comprehensive study published in 2024 analyzed outcomes for children with speech sound disorders and found that tele-therapy delivered via interactive platforms yielded comparable improvements in articulation accuracy to traditional methods [19].

However, the success of this modality depends heavily on the "ecological validity" of the intervention, monitoring the child in their natural home environment provides clinicians with more representative data than clinical snapshots. This reinforces the need for platforms that include an "Admin Interface" for therapists to review performance metrics asynchronously, a feature central to proposed systems like LogoHra.

In the domain of hearing disorders, "Tele-Audiology" has revolutionized patient management. Modern hearing aids and cochlear implants now support remote programming, allowing audiologists to adjust frequency gain and compression parameters over the cloud without requiring a hospital visit. A 2023 randomized controlled trial demonstrated that synchronous remote fine-tuning of hearing aids resulted in patient satisfaction and speech intelligibility scores statistically equivalent to face-to-face fittings [19].

Furthermore, a systematic review of digital technologies for

cochlear implant users highlighted that remote monitoring apps significantly improve patient adherence by tracking usage time and listening environments in real-time [20].

Despite these benefits, challenges remain regarding the standardization of remote diagnostics. Calibrated hardware is often required to ensure that the audio signal received by the therapist is of diagnostic quality. Recent research suggests that future mobile applications must implement "active noise monitoring" algorithms to validate the user's environment before allowing a remote hearing test or voice recording to proceed [20].

III. Problem Identification and Future Directions

To address these challenges, the forthcoming dissertation research aims to design, implement, and clinically validate a modular ecosystem for speech and hearing rehabilitation. Building upon the LogoHra platform, the specific technical objectives include:

Development of an Annotated Slovak Pediatric Dataset and ASR Fine-tuning: Curating a specialized acoustic database of dyslalic speech to fine-tune transformer-based architectures (e.g., Wav2Vec 2.0). The primary technical goal is to achieve reliable automated detection and classification of specific phonological errors (such as rotacism and sigmatism) in morphologically rich languages, robust against typical home environment noise.

Implementation of Dynamic Difficulty Adjustment (DDA) Algorithms: Developing a DDA module that utilizes real-time gameplay telemetry (e.g., ASR confidence scores, reaction times) to automatically modulate therapeutic parameters. For hearing rehabilitation, this involves dynamically adjusting the Signal-to-Noise Ratio (SNR) and auditory distractors; for speech therapy, it involves scaling the lexical and phonetic complexity of the prompted words to keep the patient in the optimal "flow channel."

Integration of an Explainable "Human-in-the-Loop" Dashboard: Creating a clinical web interface that translates raw AI data into transparent, objective metrics (e.g., automated phonological error rates over time). This addresses the clinical disconnect by ensuring the AI acts as a transparent decision-support tool rather than a "black box," allowing therapists to asynchronously configure remote therapy based on longitudinal data.

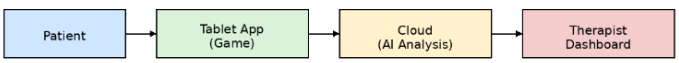


Fig. 4. Diagram of future functionality for LogoHra.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the projects KEGA 049TUKE-4/2024, by the Slovak Research and Development Agency under the projects APVV-22-0261 and APVV-22-0414.

References

[1] S. Saeedi, H. Bouraghi, M.-S. Seifpanahi, and M. Ghazisaeedi, "Application of digital games for speech therapy in children: A systematic review of features and challenges," Journal of Healthcare Engineering, vol. 2022, Article ID 4814945, 2022.

[2] H.-G. Hirsch, C. Neumann, Y. Tiggelkamp, R. Fiorista, S. Knecht, A. Schnitzler, K. Biermann-Ruben, D. Bothe, G. Bleimann, and H. Frieg, "REHALINGO – Towards a speech training system for aphasia," in *Proceedings of the 34. Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*, 2023, pp. 134–141

[3] J. R. Green, "Artificial Intelligence in Communication Sciences and Disorders: Introduction to the Forum," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 11, pp. 4157–4161, Nov. 2024.

[4] J. M. Paduthala, N. Joseph, S. Thomas, and N. A. Varghese, "Comparative Analysis of Machine Learning Models for Speech Disorder Detection," in *Proc. 2025 2nd Int. Conf. Trends Eng. Syst. Technol. (ICTEST)*, Apr. 2025.

[5] D. Lalit and J. Regash, "Explainable Artificial Intelligence in Speech and Audio Processing: Enhancing Interpretability, Fostering Trust, and Addressing Deployment Challenges," *National Journal of Speech and Audio Signal Processing*, pp. 46–53, 2025.

[6] C. Deka and A. Shrivastava, "Design and Development of a Human-Centered AI-Based Speech Therapy Tool for Children with Speech Sound Disorder," in *Intelligent Human Computer Interaction (IHCI 2024)*.

[7] J. Kim, H. Maathuis, and D. Sent, "Human-centered evaluation of explainable AI applications: a systematic review," *Frontiers in Artificial Intelligence*, vol. 7, 17 Oct. 2024.

[8] T. M. Fleming et al., "Serious Games and Gamification for Mental Health: Current Status and Promising Directions," *Frontiers in Psychiatry*, vol. 7, art. 215, Jan. 2017.

[9] J. Kováč and S. Ondáš, "LogoHra: A Customizable Gamified Speech Therapy Platform," in *2025 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2025.

[10] S. Ondáš, Z. Sokolová, J. Staš, A. Božek, "Gamification of therapeutic tools for speech and hearing disorders," in *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, IEEE, 2024.

[11] R. Senthilkumar et al., "Voice Unlocked: Transforming Speech Therapy for Special Children with VR and Gesture Animation," *Springer Proceedings*, 2025.

[12] H. Elrakaybi *et al.*, "A Machine Learning Dyslalia Mispronunciation Detection Framework for Interactive Speech Therapy," in *2025 Int. Mobile, Intell., and Ubiquitous Comput. Conf. (MIUCC)*, 2025.

[13] K. Rasheva-Yordanova et al., "A Comprehensive Review and Analysis of Virtual Reality Scenarios in Speech and Language Therapy," *TEM Journal*, vol. 14, no. 2, pp. 1895–1907, 2025.

[14] A. M. Chaves Cano, "Augmented Reality Interventions in Speech Therapy for Autism Spectrum Disorder," *Neurodivergences*, vol. 4, 2025.

[15] "Comparison of augmented reality glasses for the assistive communication support of hearing loss," *Frontiers in Neurology*, 2025.

[16] D. Modlin and Y.-T. Kuo, "The effects of using augmented reality in rehabilitation and recovery exercise on patients' outcomes and experiences: a systematic review," *Frontiers in Virtual Reality*, vol. 6, 2025.

[17] A. M. Irace et al., "Smartphone-Based Applications to Detect Hearing Loss: A Review of Current Technology," *Otolaryngology–Head and Neck Surgery*, vol. 170, no. 2, 2024.

[18] B. Göknur and M. C. Merver, "Mobile game-based auditory training as an adjunct to aural rehabilitation in preschool children with cochlear implants," *International Journal of Audiology*, 2026.

[19] T. Afouras et al., "Deep Audio-Visual Speech Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8717–8727, Dec. 2022.

[20] E. Farmani, F. Fekar Gharamaleki, and M. A. Nazari, "Challenges and opportunities of tele-speech therapy: Before and during the COVID-19 pandemic," *Journal of Public Health Research*, vol. 13, no. 1, art. 22799036231222115, Feb. 2024.

[21] M. Malmberg and J. Hagberg, "Synchronous remote fine-tuning and follow-up within aural rehabilitation: a randomised controlled trial," *International Journal of Audiology*, vol. 63, no. 6, pp. 458–466, Jun. 2024

[22] E. Laird, et al, "Systematic review of patient and service outcome measures of remote digital technologies for cochlear implant and hearing aid users," *Frontiers in Audiology and Otology*, May 2024.

Multimodality of Transformer-Based Models for Toxicity Detection in Online Games and Social Networks

[1]*Ing. Peter DEMETER (1st year)*
Supervisor: [2]prof. Ing. Kristína MACHOVÁ, PhD.

[1,2]Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovak Republic

[1]peter.demeter.2@tuke.sk, [2]kristina.machova@tuke.sk

***Abstract*—Online toxicity in multiplayer games and social networks represents a major social, ethical, and technical challenge, affecting user well-being, platform reputation, regulatory compliance, and long-term community sustainability. The scale of modern online platforms, often serving millions of users simultaneously, makes manual moderation infeasible and necessitates automated intelligent systems. Recent advances in Transformer-based architectures have significantly improved automatic toxicity detection, particularly in text-based settings, where contextual language modeling has enabled substantial performance gains over traditional approaches.**

However, contemporary online interactions are inherently multimodal, combining text, voice communication, images, short-form videos, emojis, gameplay telemetry, and social interaction patterns. Toxic behavior frequently emerges from the interaction of multiple modalities rather than from a single source of information. Despite promising progress in multimodal machine learning, unified Transformer-based frameworks specifically tailored for toxicity detection in gaming and social media environments remain underexplored.

This review summarizes the current state of research on multimodal Transformer models for toxicity detection, critically analyzes existing approaches, and identifies open challenges. The discussion highlights architectural trends, dataset limitations, fairness and robustness issues, evaluation challenges, and deployment constraints in real-world gaming and social media environments. Finally, the paper outlines potential research directions aimed at developing unified multimodal Transformer architectures capable of real-time, fair, and context-aware moderation.

***Keywords*— multimodal learning, online games, social networks, toxicity detection, transformer models**

I. Introduction

Online multiplayer games and social networking platforms have evolved into complex digital ecosystems where users communicate, collaborate, and compete in real time. Platforms such as competitive multiplayer games, streaming services, and large-scale social media networks serve as primary communication channels for younger generations. Alongside their social and economic benefits, these platforms increasingly struggle with toxic behaviors such as hate speech, harassment, cyberbullying, griefing, discriminatory language, and coordinated abuse.

Toxicity negatively impacts user engagement, psychological well-being, and retention rates. Studies have shown that exposure to persistent online harassment can significantly affect mental health, particularly among younger users. From an economic perspective, toxicity undermines player retention and monetization strategies in gaming environments. Additionally, regulatory frameworks such as the Digital Services Act (DSA) in the European Union increase pressure on platforms to implement reliable content moderation mechanisms.

Early toxicity detection methods relied on rule-based systems, keyword filtering, and classical machine learning algorithms using handcrafted features such as TF-IDF vectors. Although computationally efficient, these approaches lacked contextual understanding and were easily circumvented through creative spelling variations or implicit language. The emergence of deep learning and, particularly, Transformer-based architectures such as BERT [1] and RoBERTa [2], significantly improved performance in text classification tasks by leveraging contextualized language representations.

Nevertheless, communication in modern online ecosystems is no longer limited to plain text. In multiplayer games, toxicity frequently appears in voice chat, in-game actions, player names, or combinations of visual and textual elements (e.g., memes). Social networks integrate text, images, emojis, GIFs, and short-form videos, where toxic intent may only be detectable through cross-modal interpretation. Consequently, unimodal text-based approaches fail to capture the full spectrum of harmful behavior.

Multimodal Transformer architectures, such as VisualBERT [3], ViLBERT [4], and CLIP [5], provide a promising direction by learning joint representations across modalities. However, their application to toxicity detection—especially in gaming environments—remains relatively limited. This paper reviews the current state of multimodal Transformer-based toxicity detection, critically evaluates existing approaches, and outlines research opportunities aligned with the focus of my doctoral research.

II. Literature Review and Analysis

A. Transformer Architectures for Text-Based Toxicity Detection

The introduction of the Transformer architecture [6] marked a paradigm shift in natural language processing (NLP). Unlike

recurrent neural networks, Transformers rely entirely on self-attention mechanisms, enabling efficient parallelization and improved modeling of long-range dependencies. Pretrained language models such as BERT [1], RoBERTa [2], and DeBERTa [7] achieved state-of-the-art results in a wide range of text classification tasks, including sentiment analysis, misinformation detection, and toxicity classification.

Large publicly available datasets, such as the Jigsaw Toxic Comment Classification Challenge [8], Hatebase, and Davidson's hate speech dataset [9], enabled systematic benchmarking and rapid progress. Fine-tuned BERT-like models consistently outperform earlier convolutional and recurrent architectures due to their ability to capture contextual semantics and subtle linguistic cues.

However, despite strong benchmark performance, several limitations remain. First, these models are highly sensitive to dataset bias. Empirical studies demonstrate that models trained on Jigsaw over-predict toxicity for identity-related terms even in neutral contexts [10]. This phenomenon indicates that high accuracy does not necessarily imply fairness or robustness.

Second, text-only models struggle with implicit toxicity, sarcasm, humor, and contextual dependency across conversation threads. A single message may appear neutral in isolation but become harmful within a broader conversational context. Most benchmark datasets treat samples independently, ignoring dialogue history.

Third, text-based Transformers are insufficient in environments where communication is partially or predominantly non-textual. In competitive online games, voice chat often dominates communication, and behavioral actions can signal toxic intent even without explicit language. Therefore, while text-based Transformers are mature and effective, they represent only a partial solution in multimodal ecosystems.

B. Multimodal Transformer Architectures

To address cross-modal understanding, multimodal Transformers extend the self-attention mechanism to jointly process information from multiple modalities. Early fusion models such as VisualBERT [3] and ViLBERT [4] align visual region features extracted from convolutional neural networks with textual token embeddings through cross-attention layers. These architectures demonstrated that joint multimodal pretraining improves performance on tasks requiring integrated reasoning.

UNITER [11] introduced improved cross-modal pretraining objectives, including image-text matching and masked multimodal modeling. CLIP [5] proposed a contrastive pretraining framework trained on large-scale image–text pairs, enabling strong zero-shot performance across vision-language tasks. These models illustrate the scalability and flexibility of Transformer-based multimodal learning.

In toxicity detection, multimodal approaches are particularly relevant for meme classification. The Facebook Hateful Memes Challenge [12] demonstrated that unimodal models fail when toxic meaning arises from the interaction between neutral images and textual overlays. Multimodal Transformers significantly outperform unimodal baselines by capturing cross-modal dependencies.

Despite promising results, multimodal toxicity detection remains underexplored compared to general multimodal reasoning tasks. Many studies simply adapt generic vision-language models without incorporating task-specific inductive biases relevant to toxic behavior. Furthermore, multimodal models require large annotated datasets and high computational resources, limiting their practical adoption.

An additional challenge lies in modality imbalance. In real-world scenarios, certain modalities may be missing or noisy. For example, speech recognition errors can distort textual transcripts, and image quality may vary. Designing robust multimodal fusion strategies capable of handling incomplete inputs remains an open research problem.

III. TOXICITY IN ONLINE GAMES

Online games introduce additional complexity beyond static multimodal classification. Communication is real-time, dynamic, and highly contextual. Voice chat, player actions, spatial positioning, and interaction history contribute to the overall interpretation of behavior.

Voice toxicity detection often relies on automatic speech recognition (ASR) followed by text classification or direct acoustic modeling [13]. Transformer-based speech representation models such as wav2vec 2.0 [14] enable effective self-supervised learning from raw audio signals. However, ASR-to-text pipelines introduce error propagation, particularly in noisy gaming environments. End-to-end multimodal models that jointly integrate acoustic embeddings, linguistic content, and contextual metadata remain insufficiently studied.

Behavioral toxicity detection leverages gameplay metrics such as team damage, inactivity, reporting frequency, and social interaction graphs [15]. Sequence models and graph neural networks have been applied to identify disruptive players. However, these approaches are often separated from language-based models.

A unified Transformer architecture capable of integrating textual, acoustic, visual, and behavioral modalities could significantly enhance contextual understanding. Toxic behavior in competitive games often unfolds across time, involving escalation patterns. Therefore, temporal modeling is essential for distinguishing between isolated emotional reactions and systematic harassment.

IV. BIAS, FAIRNESS, AND ROBUSTNESS

Transformer-based toxicity detectors are vulnerable to societal bias, adversarial manipulation, and domain shift [10]. Multimodal systems introduce additional fairness challenges, as visual and speech datasets may encode demographic biases related to accent, gender, or ethnicity.

Adversarial attacks, such as character obfuscation, meme perturbations, code-switching, or phonetic spelling variations, can significantly reduce detection accuracy [16]. While multimodal systems may increase robustness through cross-modal validation, they also increase system complexity and potential vulnerability surfaces.

Fairness should therefore be treated as a core design principle rather than a secondary evaluation metric. False positives may lead to unjust penalties, reduced trust in moderation systems, and negative user experiences. In regulated environments, transparency and explainability become critical requirements.

V. LARGE LANGUAGE AND MULTIMODAL MODELS

Recent advances in large language models (LLMs) and large

multimodal models (LMMs) demonstrate emergent reasoning capabilities, contextual generalization, and strong zero-shot performance across diverse tasks. Prompt-based toxicity detection with instruction-tuned models has shown that harmful content can often be identified without explicit task-specific training [17]. Such models leverage large-scale pretraining on heterogeneous corpora, enabling implicit understanding of nuanced language, sarcasm, and context-dependent expressions.

However, the application of LLMs in toxicity detection raises several practical and theoretical questions. First, while zero-shot performance may appear competitive, calibration and consistency across domains remain unstable. Models may exhibit unpredictable sensitivity to prompt phrasing, leading to variability in moderation outcomes. Second, LLM-based moderation systems introduce explainability challenges. Although chain-of-thought reasoning can provide textual justifications, these explanations may not reflect the actual internal decision process.

In multimodal settings, large multimodal models (e.g., vision-language instruction-tuned architectures) extend reasoning capabilities to image-text interactions. These models can interpret memes, screenshots, and gameplay captures in combination with textual metadata. Nevertheless, their computational footprint is substantial. Deploying such models in latency-sensitive environments, such as competitive multiplayer games, remains problematic without aggressive optimization.

Another concern involves controllability and alignment. Large pretrained models may inherit biases or undesirable behaviors from web-scale data. Fine-tuning for moderation tasks requires careful curation and alignment strategies to ensure compliance with platform policies and regional regulations.

Therefore, instead of directly replacing task-specific models, LLMs and LMMs may serve as high-capacity teacher models in knowledge distillation pipelines. A practical architecture could involve offline large-model supervision combined with lightweight student models optimized for real-time inference. This hierarchical moderation strategy balances generalization power and deployment efficiency.

VI. PROBLEM IDENTIFICATION AND FUTURE DIRECTIONS

A. Problems Addressed by Current Research

Over the past decade, substantial progress has been achieved in automatic toxicity detection, particularly due to the adoption of Transformer-based architectures. One of the most significant contributions of current research lies in clearly demonstrating the superiority of Transformer models over traditional machine learning approaches. Compared to earlier methods based on bag-of-words representations, support vector machines, or recurrent neural networks, pretrained Transformer models such as BERT and its variants provide richer contextual embeddings and substantially higher classification accuracy in toxic comment detection tasks. This advancement has established Transformer-based fine-tuning as the dominant paradigm in text moderation systems.

Another major achievement concerns the development of multimodal architectures capable of integrating textual and visual information. This progress has been particularly visible in meme classification tasks, where toxicity often emerges from the interaction between neutral visual content and offensive textual overlays. Multimodal Transformers such as VisualBERT, ViLBERT, and UNITER demonstrated that cross-modal attention mechanisms significantly outperform unimodal baselines. These studies confirmed that toxicity cannot always be reliably detected from text or image alone, and that joint representation learning is essential in multimodal environments such as social networks.

Research has also made important contributions in identifying and analyzing fairness and bias issues in toxicity datasets. Multiple studies have shown that models trained on large-scale datasets may exhibit unintended bias toward identity-related terms, leading to disproportionate false-positive rates for specific demographic groups. The systematic measurement and mitigation of such biases represent a crucial step toward responsible AI deployment in moderation systems.

B. Open Challenges

Although multimodal toxicity detection has advanced considerably, current research still exhibits structural limitations. One of the most significant open problems is the absence of a unified multimodal modeling framework tailored specifically to gaming environments. Existing approaches typically address individual modalities—text, images, or speech—in isolation or combine only two modalities at a time. In contrast, online games generate complex multimodal streams including textual chat, voice communication, visual context, and gameplay telemetry. To date, there is no widely adopted Transformer-based architecture capable of jointly integrating all these modalities within a single coherent framework.

Another major challenge concerns temporal and contextual modeling. Toxicity is rarely an isolated event; rather, it often develops through sequential interactions between players. Current models frequently rely on single-message classification, ignoring long-term conversational context or behavioral history. Without temporal modeling mechanisms, systems may fail to distinguish between competitive emotional expression and sustained harassment. Long-range attention mechanisms and sequential modeling strategies remain insufficiently explored in this domain.

Dataset scarcity represents an additional limitation. While several benchmark datasets exist for text-based toxicity detection, large-scale annotated multimodal datasets—particularly those derived from online games—are scarce. Privacy regulations, ethical constraints, and the proprietary nature of gaming platforms complicate data collection and annotation. This lack of standardized datasets limits reproducibility, benchmarking, and fair comparison of proposed approaches.

Practical deployment also introduces real-time constraints. Online games and social networks require moderation systems capable of low-latency inference and high throughput. Large multimodal Transformer models are computationally demanding, and naïve deployment may not meet strict latency requirements. Model compression, pruning, and distillation techniques must therefore be considered integral components of system design.

C. Proposed Research Direction

From an architectural standpoint, the proposed unified model

will explore a hierarchical multimodal Transformer design. At the first level, modality-specific encoders (text Transformer, self-supervised speech encoder, and structured feature encoder for gameplay telemetry) will produce intermediate embeddings. At the second level, a shared cross-modal Transformer layer will perform joint attention across modalities.

To effectively model temporal dynamics, the architecture will incorporate session-level memory tokens capable of aggregating interaction history over multiple time steps. This design allows modeling escalation patterns and repeated behavioral signals rather than isolated events.

A further research objective will involve systematic ablation studies to evaluate the contribution of each modality. By progressively removing text, speech, or behavioral inputs, the relative importance and redundancy of modalities can be quantified. Such analysis is critical for designing efficient real-world systems.

Another important direction concerns uncertainty estimation. Toxicity detection systems deployed in high-stakes environments should quantify prediction confidence. Bayesian approximation techniques or Monte Carlo dropout strategies may provide uncertainty-aware moderation decisions, enabling human-in-the-loop intervention for borderline cases.

VII. EVALUATION CONSIDERATIONS AND METRICS

Evaluating multimodal toxicity detection systems presents non-trivial methodological challenges. Traditional metrics such as accuracy, precision, recall, and F1-score remain standard for binary or multi-label classification. However, toxicity detection often involves highly imbalanced datasets, where non-toxic samples significantly outnumber toxic ones. In such cases, macro-averaged F1-score and area under the precision-recall curve (AUPRC) provide more informative assessments than raw accuracy.

Beyond predictive performance, fairness-aware evaluation is essential. Group-based performance disparities should be measured across demographic attributes, identity-related terms, or language varieties. Metrics such as equalized odds difference, demographic parity gap, and false-positive rate disparity provide insight into potential unintended bias.

In gaming environments, temporal consistency is another critical evaluation dimension. A moderation system should not produce inconsistent decisions across sequential interactions involving the same player. Longitudinal stability metrics and session-level evaluation protocols could therefore complement traditional message-level benchmarks.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under Contract no. APVV–22–0414.

REFERENCES

[1] DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019). Minneapolis: Association for Computational Linguistics, 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

[2] LIU, Yinhan; OTT, Myle; GOYAL, Naman; DU, Jingfei; JOSHI, Mandar; CHEN, Danqi; LEVY, Omer; LEWIS, Mike; ZETTLEMOYER, Luke; STOYANOV, Veselin. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint. 2019, arXiv:1907.11692. Available at: https://arxiv.org/abs/1907.11692

[3] LI, Liunian Harold; YATSKAR, Mark; YIN, Da; Hsieh, Cho-Jui; CHANG, Kai-Wei. VisualBERT: A simple and performant baseline for vision and language. arXiv preprint. 2019, arXiv:1908.03557. Available at: https://arxiv.org/abs/1908.03557

[4] LU, Jiasen; BATRA, Dhruv; PARIKH, Devi; LEE, Stefan. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In: Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Vancouver: Curran Associates, 2019.

[5] RADFORD, Alec; KIM, Jong Wook; HALLACY, Chris; RAMESH, Aditya; GOH, Gabriel; AGARWAL, Sandhini; SASSON, Girish; ASKELL, Amanda; MISHKIN, Pamela; CLARK, Jack; KRUEGER, Gretchen; SUTSKEVER, Ilya. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning (ICML 2021). PMLR, 2021, pp. 8748–8763.

[6] VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NeurIPS 2017). 2017, pp. 5998–6008.

[7] HE, Pengcheng; LIU, Xiaodong; GAO, Jianfeng; CHEN, Weizhu. DeBERTa: Decoding-enhanced BERT with disentangled attention. In: International Conference on Learning Representations (ICLR 2021). 2021.

[8] JIGSAW. Toxic comment classification challenge. Kaggle competition dataset. 2018. Available at: https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

[9] DAVIDSON, Thomas; WARMSLEY, Dana; MACY, Michael; WEBER, Ingmar. Automated hate speech detection and the problem of offensive language. In: *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM 2017)*. 2017, pp. 512–515.

[10] DIXON, Lucas; LI, John; SORENSEN, Jeffrey; THAIN, Nithum; VASSERMAN, Lucy. Measuring and mitigating unintended bias in text classification. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2018). New York: ACM, 2018, pp. 67–73. DOI: 10.1145/3278721.3278729.

[11] CHEN, Yen-Chun; LI, Linjie; YU, Licheng; EL Kholy, Ahmed; AHMED, Faisal; GAN, Zhe; CHENG, Yu; LIU, Jingjing. UNITER: Learning universal image-text representations. In: *European Conference on Computer Vision (ECCV 2020)*. Cham: Springer, 2020, pp. 104–120. DOI: 10.1007/978-3-030-58577-8_7.

[12] KIELA, Douwe; LIPTON, Zachary C.; BOTTOU, Léon. The hateful memes challenge: Detecting hate speech in multimodal memes. In: Advances in Neural Information Processing Systems 33 (NeurIPS 2020). 2020, pp. 2611–2624.

[13] SCHMIDT, Anna; WIEGAND, Michael. A survey on hate speech detection using natural language processing. In: Proceedings of the 5th International Workshop on Natural Language Processing for Social Media (SocialNLP 2017). 2017, pp. 1–10. DOI: 10.18653/v1/W17-1101.

[14] BAEVSKI, Alexei; ZHOU, Yuhao; MOHAMED, Abdelrahman; AULI, Michael. wav2vec 2.0: A framework for self-supervised learning of speech representations. In: *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. 2020, pp. 12449–12460.

[15] BLACKBURN, Jeremy; KWAK, Haewoon. STFU noob!: Predicting crowdsourced decisions on toxic behavior in online games. In: *Proceedings of the 23rd International Conference on World Wide Web (WWW 2014)*. New York: ACM, 2014, pp. 877–888. DOI: 10.1145/2566486.2567987.

[16] HOSSEINI, Hossein; KANNAN, Sreeram; ZHANG, Baosen; POOVENDRAN, Radha. Deceiving Google's Perspective API built for detecting toxic comments. *arXiv preprint*. 2017, arXiv:1702.08138. Available at: https://arxiv.org/abs/1702.08138

[17] SCHICK, Timo; SCHÜTZE, Hinrich. Exploiting cloze questions for few-shot text classification and natural language inference. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*. 2021, pp. 255–269. DOI: 10.18653/v1/2021.eacl-main.

A Survey of Anomaly Detection Techniques for Distributed IT Systems

[1]*Svitlana SYZRANOVA (1st year)*
Supervisor: [2]Matus PLEVA

[1,2]Dept. of Computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]svitlana.syzranova@student.tuke.sk, [2]matus.pleva@tuke.sk

***Abstract*— Distributed IT systems constitute the backbone of modern computing infrastructures, including cloud platforms, cyber-physical systems, and large-scale service-oriented architectures. Their increasing scale, heterogeneity, and dynamic behavior significantly complicate system monitoring and cybersecurity assurance. Anomaly detection has therefore emerged as a key technique for identifying failures, performance degradation, and malicious activities in such environments. This paper presents a comprehensive survey of anomaly detection techniques for distributed IT systems, covering classical statistical methods, isolation-based approaches, and recent machine learning and deep learning solutions. Particular emphasis is placed on log-based anomaly detection, multivariate time-series analysis, and graph-based modeling, which have shown strong potential for capturing temporal, semantic, and structural characteristics of distributed systems. The surveyed literature is critically analyzed with respect to scalability, adaptability, data requirements, and explainability. Based on this analysis, the paper identifies unresolved challenges such as concept drift, lack of labeled data, and limited interpretability, and outlines feasible future research directions focused on hybrid, adaptive, and explainable anomaly detection frameworks for distributed IT systems.**

***Keywords*— Anomaly detection, distributed systems, deep learning, log analysis, time-series analysis, cybersecurity.**

I. Introduction

Distributed IT systems form the backbone of modern computing infrastructures, including cloud and IoT environments. While such systems offer scalability, flexibility, and fault tolerance, they also introduce significant challenges related to observability, reliability, and security. Their complexity generates large volumes of logs and metrics, making manual monitoring impractical. Consequently, automated anomaly detection has emerged as a critical capability for maintaining system availability and protecting against cyber threats [4,12].

Deep learning methods, including RNNs, CNNs, and Transformers, improve detection accuracy but introduce challenges such as high computational cost and limited explainability [8,13].

This paper presents a comprehensive survey of anomaly detection techniques for distributed IT systems, focusing on classical, machine learning, and deep learning approaches with particular emphasis on log-based, time-series, and graph-based methods. By critically analyzing recent research, this work identifies key strengths and limitations of existing solutions and outlines open research challenges that are feasible targets for further doctoral research in the area of anomaly detection and cybersecurity for distributed systems [1,6].

II. Current anomaly detection approaches

A. Classical and Statistical Anomaly Detection Methods

Early research on anomaly detection in distributed and large-scale systems predominantly relied on classical statistical, distance-based, and density-based techniques. These methods operate under the assumption that anomalous data points are rare and significantly different from normal observations in terms of statistical properties or geometric distance. Representative approaches include z-score analysis, k-nearest neighbor–based detection, clustering-based outlier identification, and probabilistic density estimation [4,12].

Among classical techniques, isolation-based methods have gained particular attention due to their scalability and robustness in high-dimensional settings. Isolation Forest and its extensions exploit the principle that anomalies are easier to isolate through random partitioning than normal instances. Isolation-based methods scale efficiently and work well in high-dimensional data. Recent surveys emphasize that such methods are especially effective for detecting global anomalies and are well-suited for unsupervised settings where labeled data is unavailable [4].

Despite their advantages, classical and isolation-based methods exhibit notable limitations when applied to modern distributed IT systems. First, they generally assume static data distributions and struggle to adapt to concept drift caused by evolving workloads, software updates, or infrastructure changes. Second, these methods typically operate on flattened feature representations, which prevents them from capturing temporal dependencies, causal relationships, and structural interactions inherent in distributed architectures. As a result, their detection performance degrades significantly in environments characterized by complex service dependencies and time-varying behaviour [1,4].

Furthermore, classical approaches often require careful feature engineering and threshold tuning, which limits their applicability in heterogeneous environments such as cloud-native and IoT-based systems. Surveys focusing on distributed and cyber-physical systems highlight that manually designed features and static thresholds are insufficient to handle the scale, diversity, and noise present in real-world operational data. These shortcomings have motivated a gradual transition toward learning-based methods that can automatically extract

relevant representations from raw logs, metrics, and event streams [6,7].

B. Anomaly Detection in Distributed and IoT Systems

Anomaly detection in distributed and IoT systems introduces challenges that go beyond those addressed by generic anomaly detection techniques. Such environments are characterized by large-scale deployment, heterogeneous devices, decentralized control, and strong coupling between cyber and physical components. Anomalies in IoT systems are diverse and context-dependent, arising from faults, network issues, or attacks. Consequently, anomaly detection solutions must operate under strict constraints related to latency, resource availability, and partial observability [6,7].

Early approaches in this domain largely relied on rule-based systems and lightweight statistical models deployed at the edge or gateway level. While these methods offer low overhead and fast response times, they suffer from limited adaptability and poor generalization across diverse operating conditions. Rule-based methods lack adaptability and lead to high false-positive rates [9].

To overcome these limitations, machine learning–based techniques have been increasingly adopted for distributed and IoT systems. Supervised and semi-supervised models, including support vector machines and shallow neural networks, have demonstrated improved detection accuracy by learning discriminative patterns from historical data. However, multiple surveys highlight that labeled anomaly data is scarce and costly to obtain in real-world deployments, particularly for rare failure modes and zero-day attacks. This dependency on labeled data significantly restricts the scalability and robustness of supervised approaches in distributed environments [6,12].

More recent research has shifted toward unsupervised and self-learning methods capable of operating without explicit labels. Deep learning–based frameworks leveraging autoencoders, recurrent architectures, and hybrid models have shown promise in capturing non-linear correlations across distributed components. Nevertheless, existing studies point out that many proposed solutions are evaluated under controlled experimental conditions and neglect deployment-related concerns such as communication overhead, model drift, and explainability. As a result, there remains a significant gap between research prototypes and operationally viable anomaly detection systems for large-scale distributed and IoT infrastructures [1,9].

C. Log-Based Anomaly Detection in Distributed Systems

Log-based anomaly detection has become one of the most prominent research directions for monitoring distributed IT systems, as logs provide detailed and fine-grained records of system behavior across software components and services. However, logs generated in distributed environments are typically high-volume, semi-structured, heterogeneous, and continuously evolving, which significantly complicates automated analysis [3,8].

Early log-based anomaly detection methods relied on pattern mining, frequency analysis, and simple statistical models applied to parsed log templates. These approaches require an initial log parsing phase to convert raw textual logs into structured event sequences. Surveys indicate that traditional parsers and template-based techniques are sensitive to log format changes and often fail to generalize across different systems or software versions. Consequently, errors introduced during the parsing stage propagate to the detection model, leading to reduced accuracy and increased false alarms [2,8].

To address the limitations of manual feature engineering, deep learning–based approaches have been widely adopted for log anomaly detection. Recurrent neural networks, particularly LSTM-based architectures, have been extensively used to model sequential dependencies between log events and to predict normal execution patterns. When deviations from predicted sequences occur, they are flagged as anomalies. While these methods demonstrate strong performance in capturing temporal dependencies, several studies report that purely sequential models struggle with long-range dependencies and suffer from high training costs, especially in large-scale distributed systems [2].

More recent approaches integrate convolutional neural networks and attention mechanisms to enhance feature extraction and improve scalability. CNN-based models are effective in capturing local patterns in log sequences, while attention mechanisms enable the model to focus on semantically important events and long-distance dependencies caused by remote procedure calls in distributed architectures. Research combining CNN, LSTM, and attention modules reports improved detection accuracy and reduced false-positive rates compared to single-model architectures. Nevertheless, these hybrid models often introduce additional complexity and require careful architectural tuning to remain deployable in real-time operational environments [3].

Transformer-based models have recently emerged as a powerful alternative for log representation learning due to their ability to model global context without recurrent structures. By leveraging self-attention, Transformer encoders can capture both local and long-range relationships among log events, making them well-suited for highly distributed systems with complex interaction patterns. Surveys note that Transformer-based log anomaly detection methods outperform traditional RNN-based approaches in terms of adaptability and contextual understanding. However, their high computational and memory requirements remain a major obstacle for deployment in resource-constrained environments [3,8].

Despite significant progress, log-based anomaly detection faces several unresolved challenges. First, most existing methods assume relatively stable log generation patterns, making them vulnerable to concept drift caused by software updates or configuration changes. Second, many approaches lack interpretability, providing limited insight into the root causes of detected anomalies, which reduces their usefulness for system operators. Finally, evaluation practices are inconsistent across studies, with limited use of standardized datasets and metrics, hindering objective comparison between methods. These limitations indicate that log-based anomaly detection, while highly effective, remains an open and active research area in distributed IT systems [1,8].

D. Time-Series and Multivariate Anomaly Detection

Time-series–based anomaly detection plays a central role in monitoring distributed IT systems, as many key indicators of system health – such as CPU utilization, network latency, throughput, and resource contention—are naturally represented as temporal signals. In contrast to univariate monitoring, multivariate time-series anomaly detection considers correlations across multiple metrics and components, which is essential for capturing coordinated failures and cascading effects in distributed environments. Multivariate analysis is

essential for capturing dependencies across system components [12,13].

Traditional time-series anomaly detection methods rely on forecasting or statistical modeling techniques, including autoregressive models, seasonal decomposition, and control charts. While these approaches are computationally efficient and interpretable, they assume stationary behavior and linear relationships, assumptions that rarely hold in modern distributed systems. Studies consistently report that such methods fail to adapt to abrupt workload changes, dynamic scaling, and non-linear interactions between system components, resulting in degraded performance under real operational conditions [5,13].

Deep learning–based time-series models have been proposed to overcome these limitations by learning complex temporal and cross-dimensional patterns directly from data. Autoencoders, recurrent neural networks, and hybrid architectures combining CNNs and RNNs are widely used for reconstructing or predicting normal behavior, with reconstruction or prediction errors serving as anomaly indicators. Although these models demonstrate superior detection accuracy, surveys highlight persistent challenges related to concept drift, hyperparameter sensitivity, and the difficulty of distinguishing rare but benign events from true anomalies in highly dynamic systems [5,13].

Furthermore, most existing time-series anomaly detection studies evaluate models in offline or near-static settings, which limits their applicability to real-time monitoring scenarios. In distributed systems, anomaly detection must operate continuously and adapt to evolving data distributions without frequent retraining. The lack of robust online learning mechanisms and standardized evaluation benchmarks remains a significant gap, motivating the integration of adaptive and self-learning techniques in future research [1,5].

E. Graph-Based and Structural Approaches

Graph-based anomaly detection has emerged as a promising paradigm for modeling the structural and relational aspects of distributed IT systems. In such representations, system components, services, or hosts are modeled as nodes, while communication patterns, dependencies, or data flows form the edges. Recent surveys identify graph-based modeling as a key advancement over purely temporal or feature-based approaches [1,10].

Graph neural networks (GNNs) and related deep learning techniques are increasingly used to learn node-level and graph-level representations for anomaly detection. By aggregating information from neighbouring nodes, GNN-based models can capture spatial dependencies and propagation effects that are common in distributed failures and coordinated cyber attacks. Empirical studies demonstrate that graph-based approaches outperform traditional methods in detecting complex, system-wide anomalies that involve multiple interacting components [10].

Despite their effectiveness, graph-based anomaly detection methods face notable practical limitations. Constructing and maintaining accurate system graphs requires detailed instrumentation and consistent data collection, which may not be feasible in large or rapidly evolving environments. Moreover, GNN models are computationally intensive and often lack transparency, making it difficult for system operators to interpret detection results and perform root-cause analysis. Surveys consistently point out that these challenges limit the adoption of graph-based techniques in production systems [1].

Overall, while graph-based and structural approaches represent a significant step toward holistic anomaly detection in distributed IT systems, their successful deployment requires advances in scalability, explainability, and integration with other data modalities such as logs and time-series metrics. These observations suggest that hybrid models combining structural, temporal, and semantic information may offer a more practical and robust direction for future research [10,12].

F. Summary of Anomaly Detection Techniques

To provide a clearer comparative overview of the surveyed techniques, Table I summarizes their main characteristics, advantages, limitations, and typical application scenarios.

TABLE I
COMPARISON OF ANOMALY DETECTION TECHNIQUES IN DISTRIBUTED SYSTEMS

Method	Strengths	Weaknesses	Typical Use Case
Statistical Methods (Traditional)	Simple, interpretable	Poor performance on complex data	Threshold-based monitoring
Machine Learning (Supervised, ML-based)	High accuracy with labeled data	Requires labeled datasets	Classification of known anomalies
Machine Learning (Unsupervised, ML-based)	No labeled data needed	Lower precision, harder interpretation	Unknown anomaly detection
Deep Learning (AI-based)	Handles complex patterns, high scalability	High computational cost	Large-scale distributed systems
Isolation Forest (ML-based)	Efficient, works well on high-dimensional data	Sensitive to parameter tuning	Log anomaly detection
Log-based Methods (Hybrid)	Uses real system logs	Requires preprocessing	System monitoring and debugging

III. PROBLEM IDENTIFICATION AND FUTURE RESEARCH DIRECTIONS

A. Problems Effectively Addressed by Existing Approaches

Existing research has made substantial progress in detecting anomalies in distributed IT systems under controlled and semi-controlled conditions. Classical and isolation-based methods provide efficient unsupervised detection mechanisms that scale well to large datasets and remain effective for identifying global anomalies. Similarly, deep learning–based approaches for log analysis and multivariate time-series monitoring have demonstrated strong capability in modeling complex temporal patterns and non-linear correlations, significantly improving detection accuracy compared to traditional techniques [4,13].

In addition, research on log-based and graph-based anomaly detection has addressed several long-standing challenges related to semantic understanding and structural awareness. Log-based deep learning models successfully capture execution context and sequential dependencies, while graph-based approaches explicitly model interactions among distributed components, enabling the detection of coordinated

and system-wide anomalies. These methods represent a significant step toward holistic monitoring solutions that go beyond isolated metric analysis [8,10].

B. Open Challenges and Limitations

Despite these achievements, multiple open challenges remain unresolved. One of the most critical issues is concept drift, where the statistical properties of system data change over time due to evolving workloads, software updates, or infrastructure reconfiguration. Most existing anomaly detection models assume relatively stable behavior and require costly retraining to maintain accuracy, which limits their practicality in long-lived distributed systems. Surveys consistently highlight the lack of robust, online learning mechanisms capable of adapting to such dynamic environments [5,13].

Another major limitation concerns the scarcity of labeled anomaly data. Supervised and semi-supervised approaches rely on annotated datasets that are difficult to obtain in real operational settings, especially for rare failure modes and previously unseen attacks. While unsupervised methods mitigate this issue, they often struggle to distinguish between benign behavioral changes and genuine anomalies, leading to high false-positive rates. This trade-off between sensitivity and reliability remains a fundamental challenge in anomaly detection research [6,12].

Explainability and operational trust also remain insufficiently addressed. Many deep learning–based models function as black boxes, providing limited insight into why a particular event or behavior was classified as anomalous. This lack of interpretability reduces the usefulness of anomaly detection systems for root-cause analysis and decision support, particularly in safety-critical and security-sensitive environments. Existing literature repeatedly emphasizes that explainable anomaly detection is essential for real-world adoption but remains underexplored [1,8].

C. Feasible Directions for Future Research

Based on the reviewed literature, several promising research directions emerge that are both technically relevant and feasible within a doctoral research framework. One such direction is the development of hybrid anomaly detection models that integrate log semantics, multivariate time-series behavior, and structural information derived from system graphs. By combining complementary data modalities, these models have the potential to improve robustness and reduce false positives in complex distributed environments [8,10].

Another important direction involves adaptive and self-supervised learning techniques designed to handle concept drift without extensive retraining or labeled data. Incorporating online learning, continual adaptation, and lightweight model updates could significantly enhance the long-term usability of anomaly detection systems in production. Furthermore, integrating explainability mechanisms such as attention visualization or attribution-based methods can improve operator trust and support effective incident response [1,5].

IV. Conclusion

This paper reviewed anomaly detection techniques for distributed IT systems, including statistical, machine learning, and deep learning approaches. While modern methods improve detection accuracy, challenges such as concept drift, limited data, and poor explainability remain. Future research should focus on adaptive, hybrid, and interpretable solutions for real-world deployment.

Acknowledgement

This study was funded by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the projects KEGA 049TUKE-4/2024 & VEGA 1/0685/26.

References

[1] L. Luan, L. Long and B. Kumar, “AI-driven anomaly detection in distributed systems: A scalable and sustainable monitoring framework,” in Proc. CCET, 2025, pp. 32–36, doi: 10.1109/CCET66260.2025.11199452.

[2] X. Wei, J. Wang, C.-a. Sun, D. Towey, S. Zhang, W. Zuo, Y. Yu, R. Ruan and G. Song, “Log-based anomaly detection for distributed systems: State of the art, industry experience, and open issues,” J. Softw.: Evol. Process, vol. 36, 2024, doi: 10.1002/smr.2650.

[3] F. Lai, P. Zhang, R. Cheng and P. Xu, "Distributed Systems Anomaly Detection Based on Log,"*2021 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Hangzhou, China, 2021, pp. 72-79, doi: 10.1109/ICT-DM52643.2021.9664097.

[4] Y. Cao, H. Xiang, H. Zhang, Y. Zhu, and K. Ting, “Anomaly detection based on isolation mechanisms: A survey,” Machine Intelligence Research, vol. 22, 2025, doi: 10.1007/s11633-025-1554-4.

[5] P. Moriano, S. C. Hespeler, M. Li and M. Mahbub, “Adaptive anomaly detection for identifying attacks in cyber-physical systems: A systematic literature review,” Artif. Intell. Rev., vol. 58, p. 283, 2025, doi: 10.1007/s10462-025-11292-w.

[6] A. Aparcana-Tasayco, X. Deng, and J. Park, “A systematic review of anomaly detection in IoT security: Towards quantum machine learning approach,” EPJ Quantum Technology, vol. 12, 2025, doi: 10.1140/epjqt/s40507-025-00414-6.

[7] S. Kim, K.-J. Park, and C. Lu, “A survey on network security for cyber–physical systems: From threats to resilient design,” IEEE Communications Surveys & Tutorials, vol. 24, no. 3, pp. 1534–1573, 2022, doi: 10.1109/COMST.2022.3187531.

[8] R. B. Yadav, P. S. Kumar, and S. V. Dhavale, “A survey on log anomaly detection using deep learning,” in Proc. 8th Int. Conf. Reliability, Infocom Technologies and Optimization (ICRITO), Noida, India, 2020, pp. 1215–1220, doi: 10.1109/ICRITO48877.2020.9197818.

[9] K. Dubey, I. Batra, A. Bajpai, and A. Malik, “A survey on anomaly detection in IoT networks: From classical methods to AI-powered solutions,” in Proc. Int. Conf. Communication, Computing and Internet of Things (CCICT), 2025, pp. 353–359, doi: 10.1109/CCICT65753.2025.00062.

[10] A. Danesh Pazho, G. Alinezhad Noghre, A. Purkayastha, J. Vempati, O. Martin and H. Tabkhi, “A survey of graph-based deep learning for anomaly detection in distributed systems,” IEEE Trans. Knowl. Data Eng., early access, pp. 1–20, 2023, doi: 10.1109/TKDE.2023.3282898.

[11] R. Miani, G. Bernardo, G. Cassales, H. Senger and E. Faria, “A survey of data stream-based intrusion detection systems,” IEEE Access, early access, pp. 1–1, 2025, doi: 10.1109/ACCESS.2025.3561105.

[12] C. Hazman, A. Guezzaz, S. Benkirane, M. Azrour, V. Ravi and A. Alabdulatif, “A comprehensive survey of contemporary anomaly detection methods for securing smart IoT systems,” Comput., Mater. Continua, vol. 85, pp. 301–329, 2025, doi: 10.32604/cmc.2025.064777.

[13] Wang F, Jiang Y, Zhang R, Wei A, Xie J, Pang X. A Survey of Deep Anomaly Detection in Multivariate Time Series: Taxonomy, Applications, and Directions. Sensors (Basel). 2025 Jan 1;25(1):190. doi: 10.3390/s25010190. PMID: 39796981; PMCID: PMC11723367.

[14] M. Masdari, “Improving security using SVM-based anomaly detection: Issues and challenges,” Soft Computing, vol. 24, pp. 16819–16838, 2020, doi: 10.1007/s00500-020-05373-x.

[15] F. Pittino, M. Puggl, T. Moldaschl and C. Hirschl, “Automatic anomaly detection on in-production manufacturing machines using statistical learning methods,” Sensors, vol. 20, no. 8, p. 2344, 2020, doi: 10.3390/s20082344.

Joint Split Computing and Early Exiting for Adaptive Transformer Inference in Vehicular Edge Environments

[1]*Chris ULIČNÝ (1st year),*
Supervisor: [2]*Juraj GAZDA*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]chris.ulicny@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—Achieving efficient and adaptive real-time perception in Connected Autonomous Vehicles (CAVs) remains a major challenge due to the high computational and communication demands of transformer-based vision models. As these models become central to modern autonomous driving systems, enabling their deployment in dynamic edge-assisted environments is critical for ensuring low-latency and reliable operation. In this work, we present a comprehensive review of split computing and early exiting techniques for transformer-based CAV perception, focusing on their fundamental principles, architectural design considerations, compression strategies, and intelligent offloading decision mechanisms. We analyze recent advances in model partitioning, artificial bottlenecks, adaptive inference, and learning-based control under fluctuating vehicular network conditions. The synergy between architectural adaptation and intelligent strategy selection offers significant potential for enabling efficient transformer inference in highly dynamic vehicular environments.**

***Keywords*—Autonomous Mobility, Computer Vision, Early Exiting, Split Computing, Transformers**

I. Introduction and Motivation

Connected Autonomous Vehicles (CAVs) are emerging as a cornerstone of intelligent transportation systems, enabled by continuous connectivity to nearby edge servers typically co-located with cellular base stations [1]. By leveraging vehicle-to-infrastructure communication, CAVs promise to reduce traffic congestion, improve road safety, and lower emissions through coordinated perception and control [2]. Central to these capabilities are advanced perception systems that rely on deep learning (DL) models to interpret complex driving environments in real time [3].

Recent advances in computer vision have led to the widespread adoption of Vision Transformers (ViTs) [4], which outperform convolutional neural networks (CNNs) [5] in tasks requiring global contextual reasoning, such as semantic segmentation and object boundary delineation. These properties are particularly important for autonomous driving, where accurate scene understanding directly impacts safety-critical decision making [6]. However, the superior representational power of ViTs comes at the cost of substantially increased computational complexity and memory footprint [7], rendering fully on-board execution on CAV hardware impractical in most real-world settings [8].

Edge computing has therefore been widely investigated as a means to offload compute-intensive perception workloads from CAVs to nearby Edge Computing Servers (ECSs) [9]. Existing research has primarily focused on optimizing resource allocation for offloading decisions, employing heuristic [10] or learning-based approaches [11], including temporal modeling via Long Short-Term Memory (LSTM) networks [12] and expert-guided imitation learning (IL) [13]. Despite these advances, most prior work treats perception workloads as monolithic tasks, ignoring architectural constraints that strongly influence latency and bandwidth requirements [14].

For DNNs with sequential layer structures, *Split Computing* (SC) [15] enables partitioning model execution between the CAV and the ECS, transmitting intermediate features rather than raw sensor data. Its effectiveness critically depends on the presence of model bottlenecks [16], which modern transformer architectures often lack [17]. Early Exiting (EE) [18] offers a complementary mechanism, allowing inference to terminate at intermediate layers once sufficient confidence is achieved. Crucially, the joint selection of split and exit points must adapt to rapidly changing network conditions characteristic of CAV operation [19].

II. Literature Review and Analysis

A. Edge Computing for CAV Perception

CAV perception systems increasingly rely on DNNs to process high-resolution sensor data in real time [3]. However, executing state-of-the-art models—particularly transformer-based architectures—entirely on-board remains computationally prohibitive due to limited processing capacity and energy constraints [8].

Edge computing (EC) has therefore emerged as a practical solution, enabling task offloading to nearby ECSs [9]. In vehicular contexts, decentralized edge architectures [1], multi-RAT systems [20], and federated offloading strategies [21] have been proposed to reduce latency while improving resource utilization. Despite these advances, EC faces a fundamental trade-off: while computation is offloaded, high-resolution sensor inputs must be transmitted over wireless channels subject to fluctuating bandwidth and interference [16]. Most offloading frameworks treat perception workloads as indivisible tasks, abstracting away model internals and limiting opportunities for fine-grained optimization [14].

B. Split Computing: Foundations and Formalization

To overcome the inefficiencies of full offloading, SC [15] proposes partitioning DNN execution between device and edge, transmitting intermediate feature representations Z_ℓ rather than raw inputs. Formally, based on [16], consider a DNN $\mathcal{M}(\cdot)$ with L layers partitioned at layer ℓ:

$$Z_\ell = \mathcal{M}_H(x), \quad \hat{y} = \mathcal{M}_T(Z_\ell),$$

where $\mathcal{M}_H(\cdot)$ and $\mathcal{M}_T(\cdot)$ represent the head and tail submodels on the CAV and ECS, and Z_ℓ denotes the intermediate feature tensor. Kang et al. [15] demonstrated that DNN layers exhibit substantial heterogeneity in computational cost and tensor dimensionality, making layer-wise partitioning viable for minimizing total inference latency and energy. However, wireless upload often dominates latency—especially at early layers where feature maps remain large—so SC without compression may offer limited gains. For large transformer models [8] that lack natural dimensionality bottlenecks [17], this limitation is particularly pronounced.

C. Artificial Bottlenecks and Feature Compression

To address communication overhead, recent research introduces artificial compression modules within DNN architectures [22], [23], [24]. The connected device functions as an encoder, extracting and compressing salient features prior to transmission [25]. As illustrated in Fig. 1, BottleNet [22] proposes an autoencoder-based bottleneck at the split point. Compression is achieved via channel-wise dimensionality reduction using 1×1 convolutions and spatial downsampling via strided convolutions, with ReLU activations introducing further sparsity. A decoder on the ECS restores tensor dimensions before inference continues.

Most studies evaluate compression strategies within CNN-based architectures. Transformer models, characterised by token-based representations and global self-attention, lack natural spatial hierarchies, making direct application of CNN-style compression nontrivial. SC also provides stronger privacy guarantees compared to conventional edge computing, since transmitted intermediate representations cannot easily reconstruct raw input data [16]. Table I summarizes key split computing and compression approaches.

TABLE I
COMPARISON OF SPLIT COMPUTING AND COMPRESSION APPROACHES

Method	Mechanism	Strengths	Limitations
Neurosurgeon [15]	Layer-wise profiling for split point	Per-layer cost optimization	No compression; large tensors
BottleNet [22]	Autoencoder bottleneck	Channel/spatial compression	Accuracy loss; CNN-oriented
BottleNet++ [23]	End-to-end bottleneck	Task-aware compression	Higher training complexity
BottleFit [24]	Self-distillation	High compression with accuracy retention	Not validated on transformers
SplitEE [25]	Joint SC with early exits	Reduces computation and transmission	Exit placement not optimized for ViTs

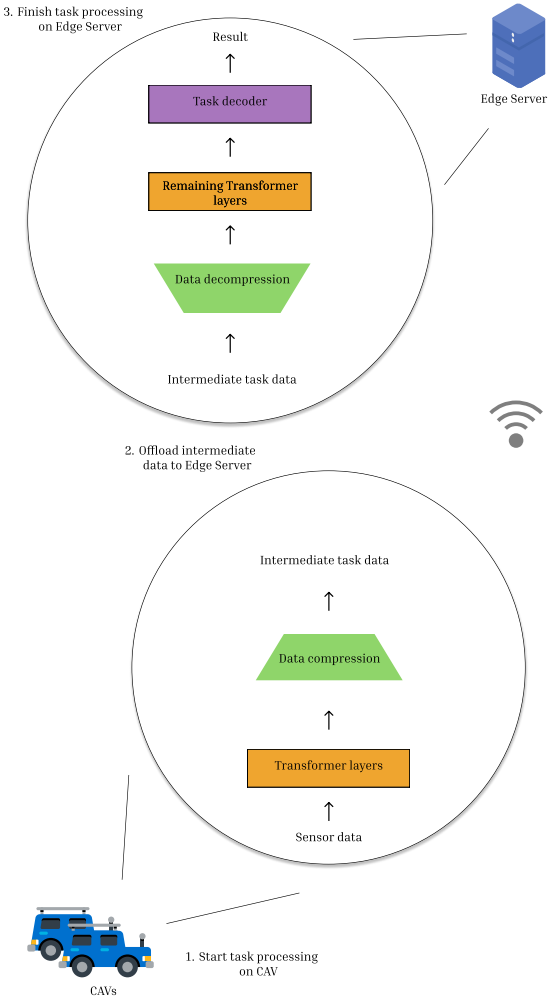


Fig. 1. Split Computing Framework with Artificial Data Compression Bottleneck within Transformer Architecture.

D. Offloading Decision Algorithms

Beyond architectural design, efficient deployment of DL models in edge environments critically depends on intelligent offloading decision algorithms [26]. These determine how and where computation should occur—on-device, fully at the ECS, or in a split configuration—balancing latency, model performance, energy, and communication overhead.

Runtime variability presents a central challenge: bandwidth fluctuates due to mobility and wireless channel conditions, device resources vary under concurrent workloads, and input characteristics change dynamically [11]. Prior research has explored heuristic approaches [10] and learning-based strategies [11]. Heuristic methods are lightweight but static; learning-based approaches improve adaptability but treat DNNs as monolithic, failing to exploit fine-grained architectural opportunities such as dynamic split points or early exits. Reinforcement learning (RL) has emerged as a promising tool for adaptive offloading [27], while LSTM-augmented [12] and IL [13] methods address temporal dynamics and training stability but have not been integrated into joint split–exit selection. Table II summarizes these paradigms.

Different split–exit combinations produce distinct computational workloads and communication requirements, necessitating dynamic adaptation. Prior studies have largely focused on CNN-based tasks, leaving transformer-based segmentation

TABLE II
COMPARISON OF OFFLOADING DECISION STRATEGIES

Strategy	Adaptability	Strengths	Limitations
Heuristic [10]	Low	Lightweight, easy to deploy	Static; poor under dynamic conditions
DL-based [11]	Medium	Learns from data	Treats DNN as monolithic
RL-based [27]	High	Online adaptation	Sample-inefficient
LSTM-augmented [12]	High	Captures temporal patterns	Not integrated with split–exit
Imitation learning [13]	Medium–High	Expert-guided stability	Bounded by expert quality

workloads insufficiently investigated. These limitations highlight the need for a unified, learning-driven framework capable of jointly selecting split–exit configurations while accounting for temporal dynamics and communication constraints.

E. Early Exiting for Adaptive Inference

Complementary to SC, EE [18] embeds auxiliary classifiers at intermediate layers, allowing inference to terminate early when sufficient confidence is achieved. Formally, based on [28], for a DNN with exit indices $C \subseteq \{1, \ldots, L-1\}$, the prediction at exit i is:

$$y_i = c_i(Z_i),$$

where Z_i represents intermediate features and c_i a lightweight prediction head. If a predefined exit criterion is satisfied, inference terminates; otherwise, computation proceeds deeper. Originally introduced in BranchyNet [18], EE was extended to ViTs [29], [30], showing that intermediate exits can reduce inference cost with minimal performance degradation. Given the high computational complexity of self-attention [31], such adaptive depth strategies are particularly valuable for transformer-based CAV perception.

Latency–Accuracy Trade-Off and Input Complexity: A central motivation for EE is the heterogeneous complexity of inputs. Simple highway scenes may require limited representational refinement, while dense urban environments demand deeper feature processing. Huang *et al.* [32] observed that allocating uniform computational resources across inputs is inefficient, while Kaya *et al.* [33] identified "overthinking," where deeper layers may override correct early predictions. EE mitigates this by enabling confident samples to exit before unnecessary computation.

Design, Placement, and Exit Policies: Designing effective multi-exit networks requires careful choices regarding exit structure, placement, and training [34]. Placement strategies range from uniform layer-based insertion [35] to FLOP-based equidistant spacing [36], though excessive exits introduce overhead and may impair convergence. Training typically relies on joint end-to-end optimization with weighted losses across exits [28], sometimes enhanced through knowledge distillation [16].

Exit policies are either static—applying fixed confidence thresholds from entropy or softmax probabilities [18], [35]—or dynamic, employing learnable controllers that optimize stopping decisions based on input features or system constraints [34]. Static policies are simple but cannot adapt to fluctuating system conditions, while dynamic policies improve adaptability at the cost of training overhead. Table III summarizes the principal approaches.

TABLE III
COMPARISON OF EARLY EXITING APPROACHES

Method	Exit Policy	Placement	Arch.	Key Limitation
BranchyNet [18]	Static	Manual	CNN	No runtime adaptability
MSDNet [32]	Static	Multi-scale	CNN	Uniform resource use
EPNet [35]	Dynamic (RL)	Uniform layer	CNN	Controller overhead
MESS [36]	Static	FLOP-equidist.	CNN	Not tested on ViTs
LGViT [30]	Dynamic	ViT-specific	ViT	Classification only
Nixon et al. [29]	Static	Encoder stages	ViT	Face recognition only

F. Analysis and Research Gap

Despite progress in split computing, early exiting, and adaptive offloading, key challenges persist for transformer-based CAV perception. As summarized in Tables I–II, existing SC and compression methods have not been validated on transformers, EE remains limited to classification on ViTs, and offloading algorithms do not jointly optimize split–exit configurations in a transformer-aware manner. Multi-stage encoder–decoder architectures introduce communication overhead unaccounted for in existing SC frameworks. CAV environments exhibit rapid fluctuations in bandwidth and on-device load [19], rendering static policies insufficient, while RL-based approaches [27] lack predictive temporal modeling for transformer workloads.

III. PROBLEM IDENTIFICATION AND FUTURE DIRECTIONS

The literature establishes strong foundations in edge-assisted inference and adaptive offloading. However, these central challenges remain:

1) **Transformer-Specific Deployment:** Encoder–decoder architectures with lateral connections introduce additional communication overhead, requiring deployment-aware structural adaptation. Compression strategies must also be tailored to transformers' global-context representations.
2) **Temporal Variability in Vehicular Environments:** Rapid fluctuations in bandwidth and computational load demand policies that can adapt proactively rather than reactively.

Addressing these challenges does not necessitate fundamental redesign of transformer backbones. Feasible research directions include:

- **Selective structural adaptation**, such as reducing communication overhead through targeted pruning of architecture components that contribute disproportionately to transmission cost.
- **Enhanced inference-time compression**, including lightweight bottlenecks and quantization techniques to

minimize data transfer without retraining large-scale models.

- **Hybrid RL–IL adaptive control**, combining reinforcement learning with imitation learning to learn a state-aware policy that selects the appropriate split–exit pair based on observed conditions, while accelerating early training and improving policy stability.
- **Predictive state modeling**, leveraging temporal models (e.g., LSTMs) to anticipate short-term network and workload variations.

A promising research direction is the development of a unified, learning-based framework capable of adapting transformer-based perception models to fluctuating vehicular network conditions while preserving accuracy and minimizing latency. Such a framework bridges architectural awareness with communication-aware deployment strategies, enabling efficient real-world CAV operation.

Acknowledgment

This work was supported by The Slovak Research and Development Agency project no. APVV SK-CZ-RD-21-0028, APVV-23-0512 and the Slovak Academy of Sciences project no. VEGA 1/0685/23.

Car SVG Vector used in Fig.1 was made by Aslan from svgrepo.com.

References

[1] M. Fardad, G.-M. Muntean, and I. Tal, "Decentralized vehicular edge computing framework for energy-efficient task coordination," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 2024, pp. 1–7.

[2] O. H. Orieno, N. L. Ndubuisi, V. I. Ilojianya, P. W. Biu, and B. Odonkor, "The future of autonomous vehicles in the US urban landscape: a review: analyzing implications for traffic, urban planning, and the environment," *Engineering Science & Technology Journal*, vol. 5, no. 1, pp. 43–64, 2024.

[3] L.-H. Wen and K.-H. Jo, "Deep learning-based perception systems for autonomous driving: A comprehensive survey," *Neurocomputing*, vol. 489, pp. 255–270, 2022.

[4] A. Singh, "Transformer-based sensor fusion for autonomous driving: A survey," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3312–3317.

[5] H. Thisanke, C. Deshan, K. Chamith, S. Seneviratne, R. Vidanaarachchi, and D. Herath, "Semantic segmentation using Vision Transformers: A survey," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106669, 2023.

[6] Q.-V. Lai-Dang, "A Survey of Vision Transformers in Autonomous Driving: Current Trends and Future Directions," *ArXiv*, vol. abs/2403.07542, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:268364135

[7] D. P. Pau and F. M. Aymone, "Mathematical Formulation of Learning and Its Computational Complexity for Transformers' Layers," *Eng*, vol. 5, no. 1, pp. 34–50, 2023.

[8] A. Ndikumana, K. K. Nguyen, and M. Cheriet, "Age of processing-aware offloading decision for autonomous vehicles in 5G open RAN environment," *IEEE Transactions on Mobile Computing*, 2023.

[9] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1697–1716, 2019.

[10] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge AI: On-Demand Accelerating Deep Neural Network Inference via Edge Computing," *arXiv preprint arXiv:1910.05316*, 2019.

[11] Z. Zhao, K. Wang, N. Ling, and G. Xing, "EdgeML: An AutoML Framework for Real-Time Deep Learning on the Edge," in *IoTDI*, 2021.

[12] W. Qi, Q. Li, Q. Song, L. Guo, and A. Jamalipour, "Extensive Edge Intelligence for Future Vehicular Networks in 6G," *IEEE Wireless Communications*, 2021.

[13] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation Learning: A Survey of Learning Methods," *ACM Computing Surveys*, vol. 50, no. 2, 2017.

[14] S. Dong *et al.*, "Task Offloading Strategies for Mobile Edge Computing: A Survey," *Computer Networks*, vol. 254, p. 110791, 2024.

[15] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J. Mars, and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," *ACM SIGARCH Computer Architecture News*, vol. 45, no. 1, pp. 615–629, 2017.

[16] Y. Matsubara, M. Levorato, and F. Restuccia, "Split computing and early exiting for deep learning applications: Survey and research challenges," *ACM Computing Surveys*, vol. 55, no. 5, 2022.

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *CoRR*, vol. abs/2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[18] S. Teerapittayanon, B. McDanel, and H. T. Kung, "BranchyNet: Fast Inference via Early Exiting from Deep Neural Networks," *arXiv preprint arXiv:1709.01686*, 2017.

[19] R. Yu and S. Li, "Exploring the associations between driving volatility and autonomous vehicle hazardous scenarios," *Accident Analysis & Prevention*, vol. 166, p. 106537, 2022.

[20] L. Bréhon, R. Kacimi, A.-L. Beylot *et al.*, "Multi-RAT-enabled edge computing for vehicle-to-everything architectures," *Ad Hoc Networks*, vol. 154, p. 103386, 2024.

[21] M. K. Hasan, N. Jahan, M. Z. A. Nazri, S. Islam, M. A. Khan, A. I. Alzahrani, N. Alalwan, and Y. Nam, "Federated learning for computational offloading and resource management of vehicular edge computing in 6G-V2X network," *IEEE Transactions on Consumer Electronics*, 2024.

[22] A. E. Eshratifar, A. Esmaili, and M. Pedram, "Bottlenet: A deep learning architecture for intelligent mobile cloud computing services," *CoRR*, vol. abs/1902.01000, 2019. [Online]. Available: http://arxiv.org/abs/1902.01000

[23] J. Shao and J. Zhang, "Bottlenet++: An end-to-end approach for feature compression in device-edge co-inference systems," *CoRR*, vol. abs/1910.14315, 2019. [Online]. Available: http://arxiv.org/abs/1910.14315

[24] Y. Matsubara, D. Callegaro, S. Singh, M. Levorato, and F. Restuccia, "Bottlefit: Learning compressed representations in deep neural networks for effective and efficient split computing," *CoRR*, vol. abs/2201.02693, 2022. [Online]. Available: https://arxiv.org/abs/2201.02693

[25] D. J. Bajpai, V. K. Trivedi, S. L. Yadav, and M. K. Hanawal, "Splitee: Early exit in deep neural networks with split computing," 2023. [Online]. Available: https://arxiv.org/abs/2309.09195

[26] T. Zheng, J. Wan, J. Zhang, C. Jiang, and G. Jia, "A survey of computation offloading in edge computing," in *2020 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 2020, pp. 1–6.

[27] D. Hortelano *et al.*, "A comprehensive survey on reinforcement-learning-based computation offloading techniques in edge computing systems," *Journal of Network and Computer Applications*, 2023.

[28] S. Scardapane, M. Scarpiniti, E. Baccarelli, and A. Uncini, "Why should we add early exits to neural networks?" *CoRR*, vol. abs/2004.12814, 2020. [Online]. Available: https://arxiv.org/abs/2004.12814

[29] S. Nixon, P. Ruiu, M. Cadoni, A. Lagorio, and M. Tistarelli, "Exploiting Face Recognizability with Early Exit Vision Transformers," in *BIOSIG*, 2023.

[30] G. Xu, J. Hao, L. Shen, H. Hu, Y. Luo, H. Lin, and J. Shen, "Lgvit: Dynamic early exiting for accelerating vision transformer," in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23. ACM, Oct. 2023, p. 9103–9114. [Online]. Available: http://dx.doi.org/10.1145/3581783.3611762

[31] Y. Tang, Y. Wang, J. Guo, Z. Tu, K. Han, H. Hu, and D. Tao, "A survey on transformer compression," 2024. [Online]. Available: https://arxiv.org/abs/2402.05964

[32] G. Huang, D. Chen, T. Li, F. Wu, L. van der Maaten, and K. Q. Weinberger, "Multi-scale dense convolutional networks for efficient prediction," *CoRR*, vol. abs/1703.09844, 2017. [Online]. Available: http://arxiv.org/abs/1703.09844

[33] Y. Kaya, S. Hong, and T. Dumitras, "Shallow-deep networks: Understanding and mitigating network overthinking," 2018. [Online]. Available: https://arxiv.org/abs/1810.07052

[34] H. Rahmath P, V. Srivastava, K. Chaurasia, R. G. Pacheco, and R. S. Couto, "Early-exit deep neural network - a comprehensive survey," *ACM Comput. Surv.*, vol. 57, no. 3, Nov. 2024. [Online]. Available: https://doi.org/10.1145/3698767

[35] X. Dai, X. Kong, and T. Guo, "Epnet: Learning to exit with flexible multi-branch network," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 235–244.

[36] A. Kouris, S. I. Venieris, S. Laskaridis, and N. D. Lane, "Multi-exit semantic segmentation networks," *CoRR*, vol. abs/2106.03527, 2021. [Online]. Available: https://arxiv.org/abs/2106.03527

Challenges and Future Directions for Explainable AI in Medical Imaging

[1]*Simona CHOVANCOVÁ (1st year),*
Supervisor: [2]*František BABIČ*

[1,2]Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovak Republic

[1]simona.chovancova@tuke.sk, [2]frantisek.babic@tuke.sk

***Abstract*—Recent advances in artificial intelligence, particularly deep learning, have significantly improved medical image analysis, including lung computed tomography (CT) and ultrasound (USG). Despite high predictive performance, these models often function as black boxes, limiting their trustworthiness and usability in clinical practice. Explainable artificial intelligence (XAI) has emerged to address this challenge by providing visual, verbal, and mathematically structured explanations that help clinicians understand model predictions. This paper reviews recent XAI approaches in medical imaging, including saliency-based methods such as LIME, SHAP, and Grad-CAM, verbal interpretability through narrative report generation, and interpretability via mathematical structure using radiomics and wavelet-based features. Case-based reasoning models such as MERA and ReX are also discussed for their ability to provide instance-level and temporal explanations. Despite these advances, challenges remain in aligning explanations with clinical reasoning and model decision processes. We identify open issues and propose future directions that provide clinicians with understandable results while examining why models make specific decisions and which features drive their predictions.**

***Keywords*—Explainable artificial intelligence (XAI), Interpretability, Health**

I. Introduction

Recent advances in artificial intelligence, particularly deep learning, have demonstrated strong performance in medical image analysis, including lung computed tomography (CT) and ultrasound (USG). However, accurate detection of lung diseases remains challenging, even for experienced clinicians, due to subtle visual patterns, operator dependency, and the presence of complex or previously unseen biomarkers. Key sonographic signs such as A-lines, B-lines, lung sliding, and pleural abnormalities are essential for diagnosing conditions including pneumothorax and hemothorax, yet their interpretation often suffers from variability and uncertainty.

Although deep learning models achieve high accuracy in lung disease detection, most operate as black boxes, limiting their applicability in clinical practice. In healthcare, where decisions directly impact patient outcomes, transparency, trustworthiness, and accountability are indispensable. Explainable artificial intelligence (XAI) has therefore emerged as a response to this challenge. According to Emanuele Neri et al. [1], XAI should be seen as a broad concept covering interpretability, explainability, transparency, justifiability, and contestability, enabling clinicians to understand and critically assess AI-driven decisions. Additionally Mienye et al. [2] emphasize that transparency involves making the AI decision process clear and understandable, trustworthiness ensures clinicians can rely on and verify AI outputs, and accountability allows users to challenge and rectify AI decisions when necessary.

It is important to distinguish between interpretability and explainability, as highlighted by Gupta et al. [3]. Interpretability refers to the degree to which a human can understand the logic behind a model's output and focuses on presenting results in a meaningful form. Explainability relates to understanding the internal mechanics of the model and how it reaches its conclusions, encompassing both interpretability and completeness. While all explainable models are interpretable, not all interpretable models are fully explainable.

The importance of these aspects has also been explicitly highlighted by Cinà et al. [4], who argue that explainability is a fundamental requirement for maintaining meaningful human oversight and trust in AI-assisted healthcare systems. In the context of lung imaging, XAI has the potential to reveal which clinically relevant patterns such as A-lines, B-lines or lung sliding drive model predictions, thereby supporting validation, bias detection and informed clinical decision-making.

II. Literature Review and Analysis

Machine learning and artificial intelligence have recently become central to medical imaging research, offering promising solutions for automated diagnosis, prognosis, and segmentation. Despite impressive performance, many deep learning models remain opaque and difficult to interpret, which limits their adoption in clinical practice. In healthcare, interpretability is critical because AI-driven decisions can directly affect patient outcomes, and the inability to understand or verify these decisions can reduce clinician trust and accountability.

To provide a structured overview, we adopt the framework proposed by Erico Tjoa et al. [5] and categorize interpretability approaches in medical imaging into three perspectives: saliency-based methods, verbal interpretability methods, interpretability via mathematical structure, and other approaches to interpretability. These categories will guide our analysis of existing methods, their applications to different medical domains, and their relevance to lung ultrasound and CT. Each perspective will be elaborated through representative methods and a general analysis of their role and use in medical imaging.

A. Saliency based methods in perceptive interpretability

Saliency based explainable AI methods are the most widely adopted approach within perceptive interpretability in medical

imaging. These techniques generate visual maps that highlight image regions contributing most to a model's prediction, enabling clinicians to assess whether the model relies on clinically meaningful structures rather than artifacts or spurious correlations. Borys et al. [6] provide a comprehensive review of saliency based XAI methods in medical imaging and emphasize their practical value in uncovering model errors, dataset biases, and shortcut learning. At the same time, they highlight a key limitation of current practice: the absence of standardized validation frameworks and clear guidelines defining what constitutes a reliable explanation for a specific clinical task.

Saliency based methods are commonly divided into perturbation based and backpropagation based approaches. Perturbation based methods such as occlusion, LIME, SHAP, or RISE are model agnostic and intuitive but may produce unrealistic inputs, which is particularly problematic for ultrasound data. Backpropagation based methods including gradient based techniques, LRP, Integrated Gradients, and CAM variants, especially Grad CAM, are computationally efficient and widely used, yet different methods may highlight inconsistent regions for the same prediction.

Naz et al. [7] use the model agnostic method LIME to generate local explanations for CNN based classification of pulmonary diseases from chest radiographs. Their ResNet50 based model is trained on the COVID CT and COVIDNet datasets, and LIME highlights image regions contributing to individual predictions. The explanations are evaluated against radiologist annotated images, showing alignment between model highlighted regions and clinically relevant areas.

Ihongbe et al. [8] compare two widely used visual XAI methods, LIME and Grad CAM, in the context of deep learning based chest radiography diagnostics. Using pretrained CNN models on open access chest X ray and CT datasets for pneumonia and COVID 19 detection, they conduct a human centered user study involving medical experts to evaluate clinical relevance, comprehensibility, and trust. While both methods were perceived as clinically meaningful, Grad CAM was generally preferred over LIME in terms of coherency and user trust, although clinicians emphasized the need for clearer explanations and better integration into clinical workflows. The study highlights that technical performance alone is insufficient and that user centered evaluation and clinician involvement are critical for successful adoption of XAI in practice.

Recent developments in brain imaging, as demonstrated by Lamba et al. [9], highlight the utility of saliency-based XAI techniques beyond traditional radiology applications. Their framework combines deep neural networks such as Xception with linear classifiers and employs Grad-CAM to generate heatmaps highlighting regions responsible for model predictions. These visual explanations provide transparency for otherwise black-box models, allowing clinicians to identify critical areas in MRI scans and supporting trustworthiness in decision-making. This example illustrates the generalizability of saliency-based XAI across different medical imaging modalities, reinforcing the importance of interpretability and transparency in clinical practice.

Kerber et al. [10] propose a deep learning based framework for automated age estimation from thorax and abdomen CT scans, aiming to derive imaging based biomarkers of biological age. A modified ResNet 18 model is trained on over 1600 non pathological CT scans and evaluated on an independent public dataset, with additional estimation of aleatoric uncertainty. For model interpretability, the authors employ Score CAM saliency maps, which consistently highlight anatomically meaningful regions such as the lumbar spine and abdominal aorta. The study demonstrates that combining saliency based explanations with uncertainty estimation can improve transparency and reliability of deep learning models in clinical imaging.

Another line of research focuses on improving the transparency of deep learning based brain tumor detection by explicitly integrating Layer-wise Relevance Propagation (LRP) into the diagnostic pipeline. In this work, Mandloi et al. [11] propose an explainable framework that combines conditional generative adversarial networks to address data imbalance with multiple pretrained CNN architectures for tumor detection and classification from brain MRI images. Synthetic MRI samples are generated using a cGAN to mitigate class imbalance and overfitting. While different pretrained models are evaluated for detection and classification performance, LRP is employed in the final stage to interpret model decisions by propagating relevance scores back to the input space. The study demonstrates that high diagnostic accuracy can be achieved without sacrificing interpretability, highlighting the potential of LRP to support transparent and clinically meaningful brain tumor analysis.

In the field of dermatology imaging, Özkurt [12] applied a convolutional neural network to the HAM10000 dataset for skin cancer detection, achieving high classification accuracy. To address the black-box nature of deep learning models, they employed SHAP and Layerwise analysis to provide both global and step-by-step explanations of the model's decisions. This approach enhances transparency in AI predictions, aligning with the principles highlighted by Neri et al. [1] and Mienye et al. [2], who stress that transparency is essential for clinicians to understand, trust, and critically assess AI-driven decisions in healthcare.

Within interpretability via mathematical structure, gradient-based explanations are commonly used with standard convolutional models. In lung ultrasonography analysis, Kolárik et al. [13] apply 3D convolutional neural networks to classify lung sliding from B-mode ultrasound videos and interpret the model using Vanilla Saliency Maps. These maps visualize pixels that most influence the CNN's predictions, offering a direct insight into the behavior of a Vanilla CNN. To improve stability and reduce noise, SmoothGrad is additionally employed, leading to clearer and more clinically meaningful highlighted regions related to pleural motion.

B. Verbal interpretability methods in perceptive interpretability

Verbal interpretability in medical AI can provide clinicians with clear, human-readable explanations of model decisions, complementing visual methods. According to Tjoa et al. [5], rule-based or textual explanations help reveal hidden patterns or unexpected correlations in the data that might otherwise go unnoticed. Despite their potential, these approaches have not yet seen widespread clinical adoption, partly due to the lack of standardized protocols and limited evidence of their usefulness in real clinical practice.

Hamamci et al. [14] explored automated report generation for 3D medical imaging to support clinical workflows

and reduce radiologist workload. Their CT2Rep framework focuses on chest CT volumes and leverages a 3D vision encoder combined with an auto-regressive causal transformer to generate structured radiology reports. The model can integrate longitudinal multimodal data using cross-attention fusion and hierarchical memory, allowing explanations to incorporate information from previous patient scans. This approach illustrates how verbal interpretability can be extended from traditional 2D imaging to more complex 3D contexts, providing clinicians with human-readable summaries alongside predictions.

Borghouts [15] developed a deep learning–based approach to generate narrative-style radiology reports from volumetric CT scans. Their encoder-decoder models are trained to identify and describe both clinically relevant abnormalities, such as pulmonary nodules and pleural effusions, and surrogate artificial abnormalities. The generated textual reports offer clinicians an interpretable explanation of the AI's assessment by highlighting detected features in natural language. This study demonstrates the potential of verbal explanations in medical imaging, providing a framework for improving transparency, trust, and usability of AI systems in clinical practice.

Moll et al. [16] explore the use of lightweight encoder-decoder models to structure radiology reports into human-readable explanations. Their approach leverages T5 and BERT2BERT models trained on the MIMIC-CXR and CheXpert Plus datasets to generate clear textual representations of imaging findings. By producing structured and comprehensible reports, these models allow clinicians to understand the rationale behind AI predictions, enhancing transparency and trust in automated medical imaging systems while remaining computationally efficient

C. Interpretability via Mathematical Structure

Interpretability via mathematical structure focuses on understanding how models make predictions by looking at their internal formulas or rules, rather than only at their outputs or visual explanations. As Tjoa et al. [5] point out, models with clear mathematical structure or predefined relationships can show how input variables affect predictions, which is especially useful in medicine. Examples include simple linear or logistic regression, parametric models based on biomedical knowledge, and neural networks combined with domain-specific rules. These approaches help clinicians see why a model made a certain prediction, making AI more trustworthy and easier to use in practice.

Interpretability in CT-based lung cancer diagnosis can be enhanced using radiomics features. Chen et al. [17] analyzed 400 ground-glass nodules from preoperative thin-section chest CT scans of 392 patients and identified eight key features with mRMR and LASSO, which allowed a transparent link between the imaging data and model predictions. Seven machine learning models were trained, with GradientBoosting performing best. The explicit use of clinically relevant features provides insight into how the model makes decisions, illustrating XAI through mathematical structure in medical imaging.

In the area of interpretability via mathematical structure, Kocadagli and Langari [18] propose an epilepsy detection framework based on electroencephalogram (EEG) signals. Their approach employs discrete wavelet transform (DWT) to decompose EEG recordings into multi-resolution frequency components, enabling the extraction of physiologically meaningful features associated with epileptic activity. These wavelet-based features provide an interpretable representation of the underlying brain dynamics and are further refined using fuzzy relations to reduce feature dimensionality. The selected features are then classified using artificial neural networks, allowing seizure-related patterns to be linked to specific frequency bands rather than opaque raw signals.

D. Other Approaches

These perspectives complement existing methods by addressing aspects such as uncertainty, robustness, and clinical applicability, while also providing alternative ways to understand and assess model behavior in medical imaging applications.

Tofighi et al. [19] introduce an explainable case-based reasoning (XCBR) framework for lung cancer diagnosis, reflecting clinicians' tendency to reason based on previously observed cases. The approach combines case-based reasoning with Naive Bayes and multilayer perceptron classifiers in a hierarchical setup, where complex or uncertain cases are further refined by the neural model. Interpretability is enhanced using SHAP values to explain feature contributions, while case complexity is explicitly incorporated into both predictions and explanations. The method is evaluated on two lung cancer datasets, achieving high accuracy and demonstrating that case-based explanations can improve transparency and trust in high-risk clinical decision support systems.

Case-based reasoning is also explored through MERA, a multimodal and multiscale self-explanatory model for lung nodule diagnosis, introduced by Lu et al. [20]. MERA integrates unsupervised and weakly supervised learning to reduce the need for annotated data and explains its decisions on multiple levels: global model-level explanations via semantic latent space clustering, instance-level case-based explanations showing similar past cases, local visual explanations through attention maps, and concept explanations based on key nodule attributes. The model was evaluated on the LIDC dataset and demonstrated high diagnostic accuracy even with only 1% annotated samples, providing transparent, clinically aligned, and robust explanations.

Anchor-based explanations focus on identifying input patterns that strongly determine a model's prediction. Building on this idea, Liu and Zhang [21] propose ReX, a framework that extends local model-agnostic explanation techniques, including Anchors, LIME, and Kernel SHAP, to handle inputs of variable lengths. ReX incorporates temporal information by modifying the sampling process and surrogate features, allowing explanations to capture how predictions evolve over time. The approach was evaluated on six models across three different tasks, demonstrating improved fidelity of explanations and better understanding of model behavior, even though the specific datasets used were not detailed in the study.

E. Summary of Reviewed Approaches

The reviewed approaches show that explainability in medical imaging can be addressed from several complementary perspectives, each with distinct strengths and limitations. Saliency-based methods are valuable for visualizing image regions that influence model predictions and are often intuitive for clinicians, but they may provide only limited insight into the actual reasoning process of the model and can vary in

stability across methods and datasets. Verbal interpretability methods, including report generation and textual explanations, improve accessibility by translating model outputs into human-readable form, yet they may remain too general or omit important details about how the prediction was formed. Interpretability via mathematical structure offers a more explicit connection between input features and model output, which can support transparency and scientific understanding. However, such approaches are often less flexible than deep learning models and may be harder to apply to highly complex imaging tasks. Other approaches, such as case-based reasoning and multimodal self-explanatory systems, provide richer and often more clinically aligned explanations by linking predictions to similar examples, semantic concepts, or temporal patterns. Their limitation is that they can be more complex to design, evaluate, and integrate into practical workflows. Overall, the reviewed methods suggest that no single explainability approach is sufficient on its own. Visual methods are effective for localization, textual methods improve communication, and structurally interpretable or case-based methods provide deeper insight into the basis of prediction. This comparison indicates that future progress in medical imaging AI will likely depend on combining these perspectives in a way that supports both clinical usability and deeper understanding of model behavior.

III. Problem Identification and Future Directions

Explainable AI in medical imaging has progressed from performance-oriented prediction toward methods that provide visual, textual, and structurally interpretable outputs. In this context, an important research focus is the development of explanations that can be directly understood and meaningfully used by clinicians in diagnostic practice. Visual outputs such as heatmaps can help localize relevant image evidence, while concise textual descriptions can present the model's findings in a more accessible and clinically interpretable form.

A central direction of future research is to examine why the model arrived at a particular prediction and which features, attributes, or internal representations contributed most strongly to its decision. This perspective is important not only for improving interpretability from the clinician's point of view, but also for enabling deeper analysis of model behavior from a scientific perspective. By combining human-understandable outputs with methods that reveal the basis of the model's reasoning, future work can contribute to AI systems that are more transparent, trustworthy, and informative for medical decision support.

References

[1] E. Neri, G. Aghakhanyan, M. Zerunian, N. Gandolfo, R. Grassi, V. Miele, A. Giovagnoni, A. Laghi, and S. E. G. on Artificial Intelligence, "Explainable ai in radiology: a white paper of the italian society of medical and interventional radiology," *La radiologia medica*, vol. 128, no. 6, pp. 755–764, 2023.

[2] I. D. Mienye, G. Obaido, N. Jere, E. Mienye, K. Aruleba, I. D. Emmanuel, and B. Ogbuokiri, "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges," *Informatics in Medicine Unlocked*, vol. 51, p. 101587, 2024.

[3] J. Gupta and K. Seeja, "A comparative study and systematic analysis of xai models and their applications in healthcare," *Archives of Computational Methods in Engineering*, vol. 31, no. 7, pp. 3977–4002, 2024.

[4] G. Cinà, T. E. Röber, R. Goedhart, and Ş. İ. Birbil, "Why we do need explainable ai for healthcare," *Diagnostic and Prognostic Research*, vol. 9, no. 1, p. 24, 2025.

[5] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 11, pp. 4793–4813, 2020.

[6] K. Borys, Y. A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C. M. Friedrich, and F. Nensa, "Explainable ai in medical imaging: An overview for clinical practitioners–beyond saliency-based xai approaches," *European journal of radiology*, vol. 162, p. 110786, 2023.

[7] Z. Naz, M. U. G. Khan, T. Saba, A. Rehman, H. Nobanee, and S. A. Bahaj, "An explainable ai-enabled framework for interpreting pulmonary diseases from chest radiographs," *Cancers*, vol. 15, no. 1, p. 314, 2023.

[8] I. E. Ihongbe, S. Fouad, T. F. Mahmoud, A. Rajasekaran, and B. Bhatia, "Evaluating explainable artificial intelligence (xai) techniques in chest radiology imaging through a human-centered lens," *Plos one*, vol. 19, no. 10, p. e0308758, 2024.

[9] K. Lamba and S. Rani, "A novel approach of brain-computer interfacing (bci) and grad-cam based explainable artificial intelligence: Use case scenario for smart healthcare," *Journal of Neuroscience Methods*, vol. 408, p. 110159, 2024.

[10] B. Kerber, T. Hepp, T. Kuestner, and S. Gatidis, "Deep learning-based age estimation from clinical computed tomography image data of the thorax and abdomen in the adult population," *Plos one*, vol. 18, no. 11, p. e0292993, 2023.

[11] S. Mandloi, M. Zuber, and R. K. Gupta, "An explainable brain tumor detection and classification model using deep learning and layer-wise relevance propagation," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 33 753–33 783, 2024.

[12] C. ÖZKURT, "Advancing skin cancer diagnosis through the comparison of shap and layer-wise relevance propagation (lrp)," 2024.

[13] M. Kolarik, M. Sarnovskỳ, and J. Paralič, "Detecting the absence of lung sliding in ultrasound videos using 3d convolutional neural networks," *Acta Polytechnica Hungarica*, vol. 20, no. 6, pp. 47–60, 2023.

[14] I. E. Hamamci, S. Er, and B. Menze, "Ct2rep: Automated radiology report generation for 3d medical imaging," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 476–486.

[15] M. Borghouts, "Automatically generating narrative-style radiology reports from volumetric ct images; a proof of concept," *arXiv preprint arXiv:2406.12599*, 2024.

[16] J. Moll, L. Fay, A. Azhar, S. Ostmeier, S. Gatidis, T. C. Lueth, C. Langlotz, and J.-B. Delbrouck, "Structuring radiology reports: Challenging llms with lightweight models," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025, pp. 7718–7735.

[17] R. Chen, H. Zhang, X. Huang, H. Han, and J. Jian, "Ct radiomics-based machine learning approach for the invasiveness of pulmonary ground-glass nodules prediction," *European Journal of Radiology Open*, vol. 15, p. 100680, 2025.

[18] O. Kocadagli and R. Langari, "Classification of eeg signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations," *Expert systems with applications*, vol. 88, pp. 419–434, 2017.

[19] A. B. Tofighi, A. Ahmadi, and H. Mosadegh, "A novel case-based reasoning system for explainable lung cancer diagnosis," *Computers in Biology and Medicine*, vol. 185, p. 109547, 2025.

[20] J. Lu, C. Yin, S. Ingala, K. Erleben, M. B. Nielsen, and S. Darkner, "Mera: Multimodal and multiscale self-explanatory model with considerably reduced annotation for lung nodule diagnosis," *arXiv preprint arXiv:2504.19357*, 2025.

[21] J. Liu and X. Zhang, "Rex: A framework for incorporating temporal information in model-agnostic local explanation techniques," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 18, 2025, pp. 18 888–18 896.

Research and Development of Artificial Intelligence for Indexing and Structuring Medical Data

[1]*Sylvia Mat'ašová (2nd year),*
Supervisor: [2]*Martin Chovanec*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]sylvia.matasova@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*—The current digital transformation in healthcare leads to an unprecedented increase in the volume of generated data. These information sets represent a heterogeneous whole, which includes not only structured database records but predominantly unstructured textual content. This paper summarizes the development of the MedGraph-AI architecture. This solution represents a hybrid neuro-symbolic approach to indexing and structuring medical data. The methodology proposes combining Deep Learning, specifically the ClinicalBERT model, with Graph Neural Networks (GNN). To ensure semantic correctness, the system integrates ontological validation. The paper provides a rigorous analysis of current NLP approaches, identifies the limits of Transformer models regarding hallucinations, and presents recent achievements in designing a robust architecture for automated entity extraction. The proposed method aims to transform passive data repositories into active knowledge sources, thereby accelerating clinical research and improving decision-making processes.**

***Keywords*—Artificial Intelligence, Medical Data Structuring, Neuro-Symbolic AI, NLP, MedGraph-AI, ClinicalBERT, Graph Neural Networks.**

I. Introduction and Motivation

The contemporary healthcare sector is undergoing a massive digital transformation, generating petabytes of patient data. As noted by Amiraslani and Dragovich [1], the definition of a document must be broad, covering historical, contemporary, and emergent forms of records. In the medical context, this heterogeneity is critical. While structured data (such as laboratory results stored in relational databases) are easily machine-readable, a significant portion of clinical knowledge remains locked in unstructured formats—primarily free-text clinical notes, discharge summaries, and operative reports.

Effective management and interpretation of this data are critical for improving healthcare quality, presenting an urgent challenge for the computer science community to transform passive repositories into active knowledge sources. In the field of applied Artificial Intelligence (AI) in medicine, research has long prioritized image analysis, particularly in radiology and diagnostics using imaging methods. Despite indisputable progress in image processing, the domain of non-image data—specifically clinical texts—remains insufficiently saturated with robust solutions. This research specifically addresses this gap by focusing on the semantic analysis of text.

The primary motivation is the limitation of current systems to effectively index and retrieve information based on context rather than mere keyword matching. The inability to structurally query unstructured notes leads to information overload for medical professionals and potential oversight of critical patient history details. Therefore, the core motivation of this work is to design a system that not only extracts information but understands the causal relationships between medical entities, as defined by the proposed MedGraph-AI architecture.

II. Background and Progress

A. Structured vs. Unstructured Data

To understand the complexity of the research task, it is essential to distinguish between data types. Structured data represents information with a high degree of organization, characterized by a clearly defined format and mutual relationships. According to Singh et al. [2], correct interpretation of the internal structure of records is a critical prerequisite for their effective retrieval and indexing. Standard relational models, while efficient for administrative data, often fail to capture the nuances of clinical narratives.

As illustrated in Fig. 1, relational tables like the OMOP CDM can store diagnosed conditions but struggle with rich free text, where the strict schema imposes a rigidity often incompatible with the natural flow of clinical documentation. This necessitates the development of systems that can bridge the gap between the flexibility of free text and the queryability of structured databases.

B. Deep Learning and NLP Limitations

In the initial research stage, we analyzed modern NLP methods. Deep Learning (DL) has transformed text processing, where architectures such as CNN and RNN have been superseded by the Transformer. Models like BERT and its medical variants (ClinicalBERT) provide state-of-the-art results in tasks like Named Entity Recognition (NER).

However, as highlighted by Kotei and Thirunavukarasu [3], while these mod-els excel at statistical pattern matching, they function as "black boxes". The "black box" problem refers to the non-transparency of the model's internal processes. In a clinical setting, this is a major hurdle. Furthermore, Large Language Models (LLMs) are prone to "hallucinations"—generating plausible but factually in-correct information. Shakil et al. [4] emphasize that factual accuracy is paramount in medical summarization. A pure deep learning approach lacks the symbolic grounding to verify if a generated diagnosis matches the symptoms described in the text according to medical protocols.

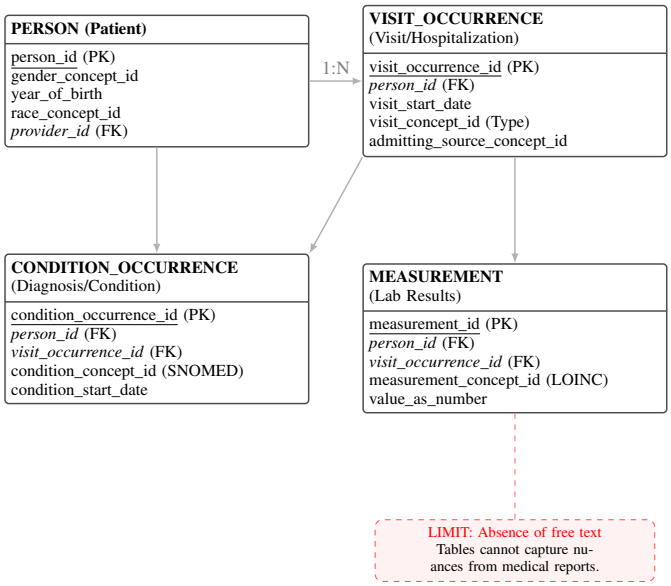


Fig. 1. Representation of structured medical data in a relational model highlighting limitations regarding free text.

C. Symbolic Approaches and Expert Systems

To address the lack of interpretability, we revisited classical Symbolic AI. Expert systems and rule-based inference mechanisms offer high interpretability but suffer from the "knowledge acquisition bottleneck". They cannot easily generalize to new, unseen text formats without manual rule updates. The architecture of such systems relies on explicit IF-THEN rules. As visualized in Fig. 2, these systems provide a clear path from user input to decision, but their rigidity makes them unsuitable for processing the vast heterogeneity of modern clinical data.

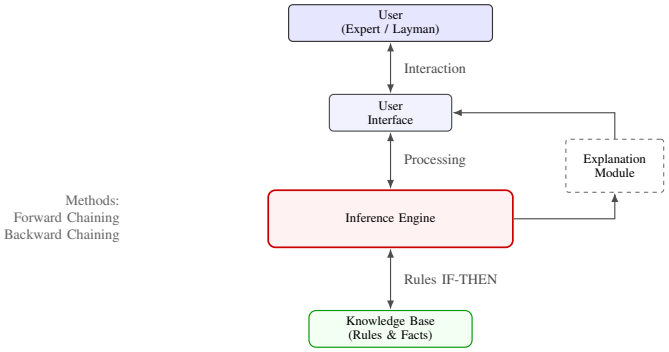


Fig. 2. Architecture of a Symbolic Expert System providing interpretability.

III. Recent Achievements

A. Proposed Methodology: MedGraph-AI

Based on the analysis of the current state, we have formulated a new methodology designated as **MedGraph-AI**. This architecture innovatively combines feature extraction using deep neural networks with semantic validation via medical ontologies. The core achievement is the design of a hybrid system where the probabilistic outputs of a neural network (such as ClinicalBERT) are not taken as final truth but are corrected by strict logical rules from a knowledge base. This minimizes the risk of hallucinations. The system utilizes Graph Neural Networks (GNN) to model the semantic context and relationships between entities extracted from the text.

The architecture is designed to operate in three distinct phases:

1) **Ingestion and Preprocessing:** Utilizing Metadata Parsers and Semantic Extractors (ClinicalBERT) to process unstructured documentation.
2) **Structuring and Indexing:** Constructing a graph from vector representations and utilizing GNN for relation inference.
3) **Ontological Validation:** Using SNOMED CT to validate extracted relations and outputting to HL7 FHIR standards.

Fig. 3 illustrates the high-level architecture of the proposed system, highlighting the interplay between the neural and symbolic components, specifically the error correction feedback loop.

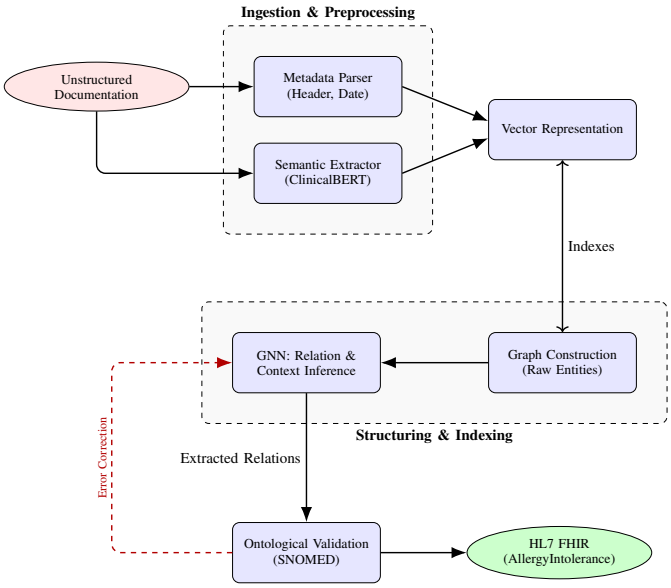


Fig. 3. Architecture of MedGraph-AI with HL7 FHIR output integration.

B. Performance Analysis and Algorithm Selection

We conducted a comprehensive structured comparison of AI algorithms for processing unstructured medical texts. The selection of the optimal algorithm depends on the specific task requirements.

Our experiments and theoretical analysis show that while Transformers (BERT) achieve the highest accuracy (F1-score 0.91 – 0.95), they suffer from critical data requirements and very low interpretability. In contrast, Deep Learning models like CNNs offer a balance but still lack transparency. Hybrid approaches, such as the one proposed in MedGraph-AI, aim to combine the high accuracy of Transformers with the interpretability of symbolic systems.

Table I summarizes our findings regarding the performance trade-offs of different algorithmic approaches in the clinical domain.

C. Expected Results and Scalability

We have modeled the expected performance of MedGraph-AI against conventional NLP methods. Scalability is a key advantage of AI-driven processing. Cloud solutions and distributed algorithms allow for the efficient processing of data

TABLE I
COMPARISON OF ALGORITHM PERFORMANCE IN PROCESSING CLINICAL TEXTS

Algorithm	Accuracy (F1-score)	Interpretability	Data Requirement
SVM / Naive Bayes	0.72 – 0.78	High	Low
Deep Learning (CNN)	0.84 – 0.89	Low	High
Transformers (BERT)	0.91 – 0.95	Very Low	Critical

streams from hospital systems. Rahmani et al. [5] emphasize that such systems are practical for enterprise environments.

As illustrated in Fig. 4, we anticipate that the integration of semantic validation will significantly increase not only the extraction accuracy but, more importantly, the robustness when processing non-standard medical reports. While conventional NLP methods often fail when facing irregular document structures (Accuracy 81%, Robustness 76%), the proposed neuro-symbolic architecture is designed to maintain high performance (Accuracy >94%, Robustness >89%).

The robustness metric here is defined as the system's ability to maintain performance in the presence of noisy data, such as typos, abbreviations, and non-standard grammatical structures commonly found in doctor's notes. This resilience is a direct result of the symbolic layer, which can infer the correct entity even if the neural network is uncertain, by relying on the knowledge base constraints. The reduction in manual validation time is supported by findings from Baviskar et al. [6], who noted up to 95% time savings with AI automation.

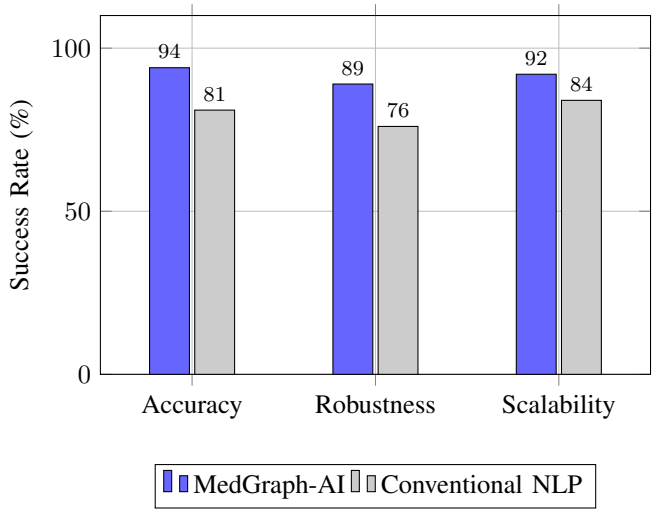


Fig. 4. Projected comparison of the proposed architecture's performance (Target metrics) against conventional methods.

D. Optimization Strategies

We also addressed the computational complexity inherent in processing high-dimensional data. Hardware acceleration using GPUs and Edge processing are essential for reducing latency in real-time document processing [7].

The recent achievements also include the theoretical formulation of the feedback loop mechanism, where error signals from the ontological check serve to adapt the neural model. This "Human-in-the-loop" concept is crucial for building trust in AI systems within the medical community.

IV. FUTURE PLANS

Based on the established theoretical framework, future work is clearly defined. The primary goal is the full implementation of the MedGraph-AI architecture.

Specific steps include:

- **Implementation of the Feedback Mechanism:** We will implement the module where error signals from the ontological validation layer are back-propagated to the ClinicalBERT model to fine-tune its embeddings. This requires developing a custom loss function that incorporates both the standard cross-entropy loss from the NLP task and a penalty term derived from ontological inconsistencies.
- **Dataset Creation:** We plan to create a specialized testing dataset from clinical data (anonymized) to benchmark the system's ability to detect latent relationships between entities that are not explicitly stated in the text. This dataset will focus on "long-tail" diseases which are often underrepresented in standard training sets.
- **Evaluation:** Realization of experiments focusing on the precision of relationship extraction compared to purely statistical methods. We aim to verify the thesis that neuro-symbolic integration significantly reduces hallucinations. Metrics will include Precision, Recall, F1-score, and a custom "Clinical Validity Score" assessed by medical experts.
- **Comparative Analysis and Ablation Studies:** Beyond standard metrics, we will conduct ablation studies to quantify the contribution of individual components of the architecture. We also plan to compare the performance of MedGraph-AI with current Large Language Models (LLMs) in a zero-shot setting to demonstrate the advantages of a specialized neuro-symbolic approach.

V. CONCLUSION

This paper presented the progress in the research of AI methods for structuring medical data. By identifying the limitations of current Deep Learning models and proposing the MedGraph-AI neuro-symbolic architecture, we aim to bridge the gap between high-performance statistical models and the need for reliable, interpretable clinical knowledge management. The integration of graph neural networks with established medical ontologies offers a promising path toward solving the "black box" problem in medical AI. Future work will focus on the empirical validation of these concepts through rigorous testing on real-world clinical datasets.

REFERENCES

[1] F. Amiraslani and D. Dragovich, "A review of documentation: A cross-disciplinary perspective," *World*, vol. 3, no. 1, pp. 126–145, 2022.
[2] A. Singh *et al.*, "Review of indexing techniques in information retrieval," *International Journal of Recent Technology and Engineering (IJRTE)*, 2022.
[3] E. Kotei and R. Thirunavukarasu, "A systematic review of transformer-based pre-trained language models through self-supervised learning," *Information*, vol. 14, no. 3, 2023.
[4] H. Shakil, A. M. Mahi, P. Nguyen, Z. Ortiz, and M. T. Mardini, "Evaluating text summaries generated by large language models using openai's gpt," 2024.
[5] A. M. Rahmani *et al.*, "Artificial intelligence in healthcare: A comprehensive review," *Journal of King Saud University - Computer and Information Sciences*, 2021.
[6] D. Baviskar *et al.*, "Automation in medical record processing," *International Journal of Scientific Research*, 2021.
[7] C. Surianarayanan *et al.*, "Edge ai for real-time document processing," *Journal of Edge Computing*, 2023.

Temperature and Pressure Effects on Breakdown in Transformer Oil-based Nanofluids

[1]*Matej VAĽKO (1st year)*
Supervisor: [2]Juraj KURIMSKÝ

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]matej.valko@tuke.sk, [2]juraj.kurimsky@tuke.sk

***Abstract*—Dielectric Breakdown of the power transformer insulation system is proven as their predominant mode of failure. Thermal and pressure fluctuations represent significant operational influences on liquid insulation in power transformers. This work evaluates the dielectric breakdown of biodegradable transformer oils and their nano-modifications exposed to changing temperature and pressure. Based on a critical review of the latest literature on failure phenomena under these environmental conditions, the study adapts several well-known research hypotheses to guide the subsequent research aimed at the performance evaluation of newly developed nanofluids.**

***Keywords*—Transformer oil, AC breakdown strength, nanofluids, nanoparticles, pressure, temperature**

I. Introduction

Mineral oil still dominates the current electric power industry all over the world, but its limitations are well known after years of experience and difficult to overcome without additional modification. Alternative insulating fluids are therefore coming to the forefront. Although they require further research, several studies to date [1], [2], [3], [4], [5], show that they could offer better parameters and higher reliability than conventional dielectric fluids. One example of these new materials are so-called *nanofluids*, which are created by enriching a pure base fluid with different types of microscopic particles measuring nanometers in size.

The important factor that must be taken into consideration is the biodegradability of used insulating fluids. Its modern variants must be in strict compliance with eco-design legislative requirements, specifically European Commission Regulation (EU) 2019/1783 [6]. This regulation, which amends the rules for small, medium, and large power transformers, sets stricter energy efficiency limits (known as Tier 2) from July 2021, emphasizing not only the environmental safety of the medium itself but also the overall operational efficiency of the equipment – a critical priority given that the CIGRÉ A2.37 survey citing dielectric breakdown as the predominant mode of failure [7]. The use of alternative fluids, such esters or other types of base fluids (or nanofluids based on them), must therefore lead to the achievement of EU targets for reducing the carbon footprint and increasing reliability and fire safety in energy networks.

In the context of these requirements, the presented paper provides a review focusing on the analysis of the influence of pressure and temperature on the breakdown strength of various insulating fluids and concludes with research hypotheses for subsequent original research aimed at performance assessment of newly developed nanofluids.

II. The State of the Art

The importance of this research stems from the physical nature of transformer operation, during which an isochoric process takes place, where a change in temperature caused by variable equipment load and ambient conditions leads to changes in internal pressure at a constant fluid volume. For this reason, to ensure reliability, it is necessary to know the exact breakdown strength and the behaviour of other important operating parameters of insulating fluids precisely during dynamic changes in these thermodynamic conditions. The global importance of research in this area is evident from the following scientometric analysis.

A. Scientometric Analysis

Scientometrics, defined as the study of quantitative aspects of science and technology in the context of information communication by Nalimov and Mulchenko in 1971 [8], provides an analytical framework for measuring scientific performance. Its primary function is to generate relevant productivity indicators and map scientific impact, enabling not only ex-post evaluation of research activities, but also the formulation of predictions and guidelines for optimizing scientific processes.

For the purposes of our *scientometric analysis*, the *Web of Science* (WoS) citation database was chosen as the primary source due to higher quantity of relevant results compared to *Scopus* database. The search was performed using a logical product (the "AND" operator), which ensured that only those publications containing all the specified terms were included. To examine the effect of temperature, the keyword string "electrical breakdown," "nanofluids," and "temperature" was used, while to analyse the effect of pressure, due to a lack of relevant records in the original settings, a modified combination of "breakdown," "transformer oils," and "pressure" was selected, and the final data was then processed into bibliometric maps using the *VOSViewer* [9] software.

The bibliometric map in Fig. 1 visualizes the geographical distribution of research activity in the area of interest (influence of temperature on the dielectric properties of dielectric fluids), with individual countries ranked according to the total volume of published works. Analysis of mutual citation links led to the identification of four specific clusters, in which countries are ranked in descending order according

to publication performance. The largest group, the so-called red cluster, consists of India, China, Malaysia, and Iran. The second most significant, the green cluster, includes Thailand, Slovakia, and Spain, followed by the blue cluster, represented by Egypt and Greece. The yellow cluster, consisting exclusively of Saudi Arabia, has a specific position. For a more detailed quantification and direct visual comparison of the publication output of individual countries, see the supplementary bar chart in Fig. 2.

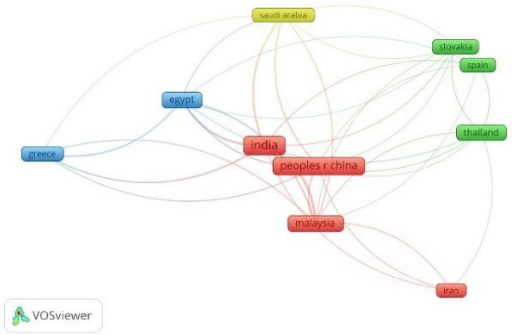


Fig. 1 Bibliometric map of countries participating in research in the area of interest (influence of temperature), based on citations and oriented according to the total number of publications

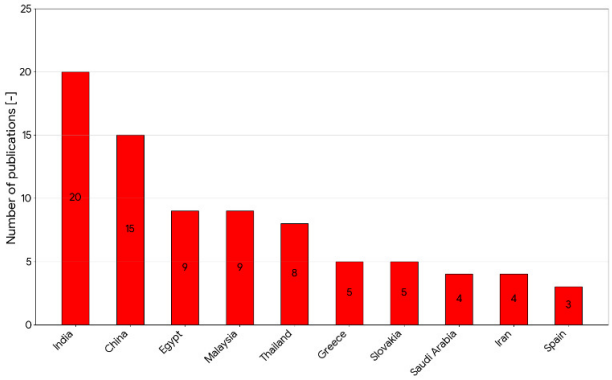


Fig. 2 Bar chart illustrating the number of publications in the area of interest (influence of temperature) by country of publication

The bibliometric map in Fig. 3 visualizes the geographical distribution of research activities in the area of interest, with the ranking of countries reflecting their overall publication performance. Based on the analysis of citation links, three evenly represented clusters and two isolated countries without direct citation links were identified. Within each group, countries are ranked in descending order by number of publications. The red cluster includes Malaysia, Canada, Italy, and Nigeria, while the green cluster consists of China, the US, Russia, and Japan. The third, blue cluster is represented mainly by European countries: France, Norway, England, and Sweden. Outside these interconnected groups are South Korea and Indonesia, which do not show any citation correlation with other entities within the analysed sample. For a more detailed quantification of the publication activity of individual countries, see the supplementary bar chart in Fig. 4.

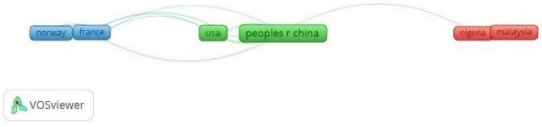


Fig. 3 Bibliometric map of countries participating in research in the area of interest (influence of pressure), based on citations and oriented according to the total number of publications

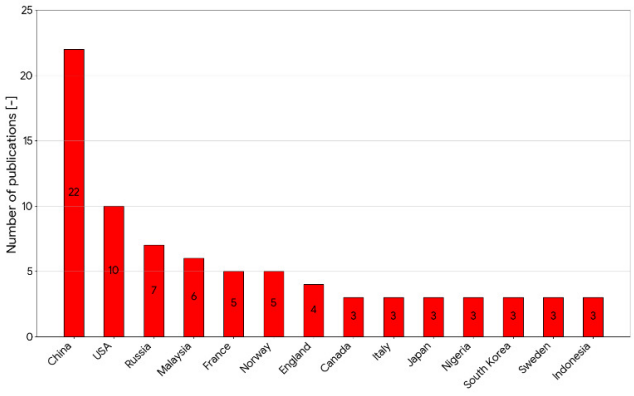
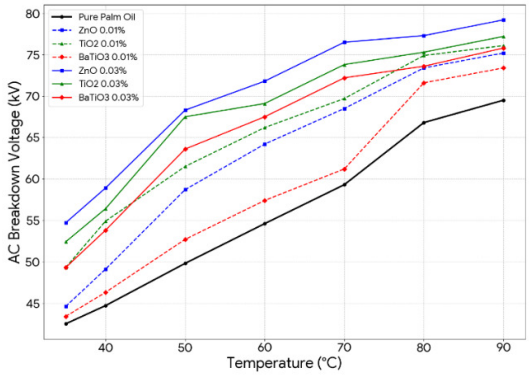


Fig. 4 Bar chart illustrating the number of publications in the area of interest (influence of temperature) by country of publication

B. Trends in the research on the influence of temperature on the breakdown of dielectric liquids

Muangpratoom's paper [10] focused on modifying the dielectric properties of refined palm oil (RPO), which is an environmentally friendly alternative for use in transformers and other HV equipment. In his work, the author examined samples with ZnO, TiO2, and BaTiO3 nanoparticles in volume concentrations of 0.01 vol.% and 0.03 vol.%. To ensure the stability and homogeneity of the resulting nanofluids, these samples were prepared using a two-step method consisting of magnetic stirring followed by ultrasonic dispersion for two hours. The experimental verification of AC breakdown strength was carried out in accordance with IEC 60156 [11] using a standardized configuration with spherical electrodes and an electrode distance of 2.5 mm, in a wide temperature range from 35 °C to 90 °C (at 35, 40, 50, 60, 70, 80, and 90 °C). The data obtained confirm that the addition of nanoparticles not only reduces the moisture content but also increases the AC breakdown voltage compared to the pure sample. This positive trend was more pronounced at higher temperatures and higher concentrations of additives, as shown in Fig. 5. The optimal mixture evaluated in the study was a nanofluid with a ZnO, temperature 90 °C, and concentration of 0.03 vol.%. This sample exhibited maximal observed AC breakdown voltage (U_{BDV} = 79.2 kV), and therefore highest dielectric strength of 31.68 kV/mm. The authors explain this phenomenon by the mechanism of capturing free electrons by nanoparticles, which effectively slows down the process of streamer discharge initiation in the liquid itself.

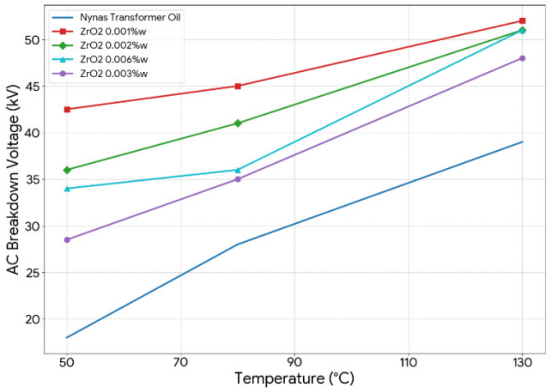


*datapoints obtained from the ref. [10]

Fig. 5 Effect of temperature on the AC Breakdown Voltage on pure palm oil and palm oil-based Nanofluids

Abd-Elhady et al. [12] analysed the effect of temperature on the breakdown voltage of nanofluids prepared by ultrasonic dispersion of ZrO_2 nanoparticles into Nynas

transformer oil (TO) at mass concentrations of 0.001, 0.002, 0.003, and 0.006 %w. Experimental measurements were performed at temperatures of 50 °C, 80 °C, and 130 °C using a mushroom electrode system (with electrode distance 2 mm) powered by a 50 Hz source with a maximum voltage of 60 kV, with 10 measurements taken at one-minute intervals for each sample. The results confirmed that the presence of ZrO_2, as shown in Fig. 6, increases the dielectric strength compared to pure oil; the maximum increase (approximately 31%) was recorded at a concentration of 0.001%w and a temperature of 130 °C (from 40 kV to 52.5 kV), while the lowest increase (25%) was recorded at a concentration of 0.003%w at 80 °C (from 28 kV to 35 kV). Despite the maximum at high temperatures, the authors note that a more significant improvement in insulating properties occurs at lower temperatures and lower concentrations, which they attribute to the ability of nanoparticles to bind charge and absorb moisture, the effectiveness of which, however, decreases with increasing temperature.



*datapoints obtained from ref. [12]

Fig. 6 Effect of temperature on the AC Breakdown Voltage on pure Nynas Transformer oil and its Nanofluids

Table I provides a systematic overview of the results of studies focusing on the temperature dependence of electrical breakdown in various types of nanofluids. From the analysis of the presented data, it can be concluded that an increase in temperature generally has a favourable effect on increasing the breakdown voltage. At the same time, however, it is confirmed that determining the optimal concentration of additives is not universal but is conditioned by the specific combination of the type of nanoparticles used and the properties of the base dielectric liquid.

TABLE I
SUMMARY TABLE OF RESEARCH ON THE EFFECT OF TEMPERATURE ON THE MAGNITUDE OF AC BREAKDOWN VOLTAGE

Base Fluid	Type of Nanoparticles	Optimal Concentration of Nanoparticles	Optimal Temperature of Sample	Maximal Improvement	Ref.
RPO	ZnO, TiO_2, $BaTiO_3$	0.03 vol.% (ZnO)	90 °C	14 %	[10]
TO	ZrO_2	0.001 %w	130 °C	31 %	[12]

C. Trends in the research on the influence of pressure on the breakdown of dielectric liquids

Norgard and Curry focused in their paper [13] on the analysis of the breakdown strength of synthetic polyalphaolefin-based (or PAO for short) insulating oil for use in high-pressure switches. Experimental verification was carried out in a pressure range of 0.5 to 10 MPa at set interelectrode distances of 0.5 mm, 0.75 mm, and 1 mm, with high data validity guaranteed by 450 repetitions for each configuration. A specialized circuit was used to generate the impulse stress, in which energy from a capacitor battery was discharged through a gas triode into the primary winding of an impulse transformer, providing a 10-times amplified voltage, which was then supplied to the switch itself. The results showed that the breakdown strength of PAO oil increases non - linearly with increasing pressure and decreasing electrode distance. The authors interpret this trend, which correlates with an increase in breakdown voltage, because of the more difficult conditions for the formation of gas bubbles in the volume of liquid under pressure, thereby eliminating the primary factor in the initiation of electrical discharge.

Kúdelčík et al. [14] focused on the experimental analysis of breakdown voltage in ITO 100 transformer oil depending on pressure in the range of 1 to 10 bar. The HT 55 – I system with a maximum output of 55 kV and a constant rise rate of 2 kV/s was used to generate the voltage, connected to spherical copper electrodes (r = 35 mm) with distances set to 1.2 mm, 1.8 mm, and 2.1 mm. The methodology included seven repetitions for each setting, with extreme values (minimum and maximum) excluded to increase accuracy, and the oil homogenized between discharges by 5 minutes of magnetic stirring. The results confirmed that the breakdown voltage increases with increasing pressure, with a maximum value of U_{MAX} = 44 kV achieved at a pressure of 7 bar and a 2.1 mm electrode distance. The authors attribute this increase to the physical phenomenon of shortening the mean free path of electrons λ in a compressed medium, which leads to a decrease in the efficiency of the ionization process and a subsequent increase in the breakdown strength of the oil.

Table II provides a systematic overview of studies that examine the dependence of electrical breakdown on pressure in various dielectric liquids. The findings suggest that, under the given experimental conditions, the dielectric strength of these media can be effectively increased by appropriate pressure regulation.

TABLE II
SUMMARY TABLE OF RESEARCH ON THE EFFECT OF PRESSURE ON THE MAGNITUDE OF AC BREAKDOWN VOLTAGE

Base Fluid	Electrode Distance	Pressure range	Maximal Improvement	Ref.
PAO	0.5, 0,75, 1 mm	0.5 – 10 MPa	12 % (0.5 mm)	[13]
ITO 100	1.2, 1.8, 2.1 mm	1 – 10 bar	65 % (2.1 mm)	[14]

III. Hypotheses and Research Outlook

As revealed by the research in Chapter II, the modification of state conditions can influence breakdown strength (breakdown voltage) not only in base dielectric liquids, but also in their nanoparticle-modified variants. This fundamental conclusion defines the direction of further research, which should expand the experimental scope not only to the incorporation of different types of samples and operating conditions, but also to the integration of advanced diagnostic and experimental methods.

A. Hypotheses

Based on the analysis of the current research direction, we present leading hypotheses that will guide the preparation of our own nano-modified transformer oils and research into their dielectric performance:

1. **Nanoparticle Composition:** The dielectric and thermal performance of insulating fluids is distinctively dependent on the classification of the nanoparticle used [10].
2. **Aging and Stability:** Nanoparticles can alter the aging processes of the base fluids. Under accelerated aging conditions (electrical, thermal or combination of both types of stresses), nanofluids will exhibit different degradation curve – either they will maintain their stability for longer than base fluid, or they will degrade faster [15].
3. **Diagnostic Sensitivity:** Results of Dielectric frequency response (DFR) analysis at specific frequencies will correlate with the ability of base fluids and their nanofluid variant to withstand electrical stress.
4. **Base Fluid Interaction:** Insulating properties of nanofluids is not universal, but is highly dependent on the type of base fluid used [16].

B. Research Outlook

The planned research will significantly expand the experimental framework beyond the current use of magnetite nanoparticles to include the application of carbon-based nanoparticles and other types of nanomaterials, enabling a more comprehensive understanding of the influence of particle type on dielectric properties. For accurate quantification of changes (improvement or degradation of parameters), parallel comparative analyses of different types of pure base oils and their corresponding nanofluids will be performed.

From a methodological point of view, the research will move from the exclusive evaluation of electrical strength (breakdown voltage) to the incorporation of DFR analysis as the diagnostic tool for sample testing in the frequency domain. This diagnostic will include measuring the dielectric losses as a function of voltage and frequency, as well as determining the relative permittivity ε_r. A comprehensive picture of the operating limits and behavior of nanoliquids under combined physical fields will be complemented by a set of multiphysical and physicochemical tests focused on measuring acidity, density, viscosity, water content, as well as flash and pour points.

Future work will focus on the long-term sample stability through accelerated aging tests. These evaluations will focus on the electrochemical and thermal degradation of the fluids, simulating the mechanisms encountered in real-world operation of power transformers. Comparative stress testing of base fluids and nanofluids is essential to the prediction of their performance in the grid applications. In addition, it is recommended to continue research on the thermodynamic dependence of liquid dielectric phase. Test protocols should consider the simultaneous effects of temperature and pressure to capture their combined effect. Understanding this interaction is crucial to the definition of the operational limits of nanofluids in the operation of power transformers.

ACKNOWLEDGMENT

This research was funded by the Ministry of Education, Research, Development and Youth within the project VEGA 2/0029/24, 1/0380/24, the Slovak Agency for Research and Development based on contracts no. APVV-22-0115.

The work also received support from REA.A – Marie Skłodowska-Curie Actions & Support to Experts within the project 101182948 — RETROTRAFO — HORIZON-MSCA-2023-SE-01.

REFERENCES

[1] H. Zhengyong *et al.*, 'Significantly enhanced electrical performances of eco-friendly dielectric liquids for harsh conditions with fullerene', *Nanomaterials*, 2019.

[2] A. Cavallini, R. Karthik, and F. Negri, 'The effect of magnetite, graphene oxide and silicone oxide nanoparticles on dielectric withstand characteristics of mineral oil', *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 5, Art. no. 5, Oct. 2015, doi: 10.1109/TDEI.2015.005016.

[3] M. Rafiq *et al.*, 'Insulating and aging properties of transformer oil-based TiO2 nanofluids', in *2014 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, Oct. 2014, pp. 457–461. doi: 10.1109/CEIDP.2014.6995798.

[4] U. Khaled and A. Beroual, 'AC Dielectric Strength of Synthetic Ester-Based Fe3O4 , Al2O3 and SiO2 Nanofluids -Conformity with Normal and Weibull Distributions', *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 26, pp. 625–633, Apr. 2019, doi: 10.1109/TDEI.2018.007759.

[5] C. Zhang, Y. Wang, Z. Yan, and Z. He, 'Effect of the Coexistence of Al2O3 Nanoparticles and Water on Transformer Oil Electrical Performance', *IEEE Access*, vol. 7, pp. 10189–10195, 2019, doi: 10.1109/ACCESS.2018.2890567.

[6] European Comission, *Commission Regulation (EU) 2019/1783 of 1 October 2019 amending Regulation (EU) No 548/2014 on implementing Directive 2009/125/EC of the European Parliament and of the Council with regard to small, medium and large power transformers*, vol. 2019/1783. 2019, pp. 107–120. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1783

[7] WG A2.37, 'Transformer reliability survey'. CIGRÉ, 2015. [Online]. Available: https://www.e-cigre.org/publications/detail/642-transformer-reliability-survey.html

[8] V. V. Nalimov and Z. M. Mulchenko, 'Measurement of Science. Study of the Development of Science as an Information Process.', Foreign Technology Div Wright-Patterson AFB Ohio, AD735634, 1971. Accessed: Mar. 04, 2025. [Online]. Available: https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/AD735634.xhtml

[9] N. J. van Eck and L. Waltman, 'VOSviewer Manual', Oct. 2023, [Online]. Available: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf

[10] P. Muangpratoom, 'The Effect of Temperature on the Electrical Characteristics of Nanofluids Based on Palm Oil', *J. Eng. Technol. Sci.*, vol. 53, no. 3, p. 210312, Jul. 2021, doi: 10.5614/j.eng.technol.sci.2021.53.3.12.

[11] International Electrotechnical Commission, *IEC 60156, Insulating Liquids-Determination of the breakdown voltage at power frequency-Test Method*, Jul. 1995.

[12] A. M. Abd-Elhady, M. E. Ibrahim, T. A. Taha, and M. A. Izzularab, 'Effect of temperature on AC breakdown of nanofilled transformer oil', *IET Journals*, vol. 12, no. 1, pp. 138–144, 2018, doi: 10.1049/iet-smt.2017.0217.

[13] P. Norgard and R. Curry, 'An Examination of the Statistics of Breakdown in a Pressurized, Flowing, Synthetic Oil Dielectric', in *2008 IEEE International Power Modulators and High-Voltage Conference*, May 2008, pp. 202–206. doi: 10.1109/IPMC.2008.4743616.

[14] J. Kúdelčík and L. Varačka, 'Pressure effects on breakdown in transformer oil ITO 100', in *2014 ELEKTRO*, May 2014, pp. 666–669. doi: 10.1109/ELEKTRO.2014.6848984.

[15] R. Karthik, F. Negri, and A. Cavallini, 'Influence of ageing on dielectric characteristics of silicone dioxide, tin oxide and ferro nanofluids based mineral oil', in *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, Feb. 2016, pp. 40–43. doi: 10.1109/AEEICB.2016.7538348.

[16] A. K. Das, 'Investigation of electrical breakdown and heat transfer properties of coconut oil-based nanofluids', *Industrial Crops and Products*, vol. 197, p. 116545, Jul. 2023, doi: 10.1016/j.indcrop.2023.116545.

Multi-Metric Patch Prioritization for Small Networks Using Attack-Graph Context

[1]*Jaroslav MARKO (2nd year),*
Supervisor: [2]*Anton BALÁŽ*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]jaroslav.marko@tuke.sk, [2]anton.balaz@tuke.sk

***Abstract*—Small networks have limited capacity to deploy fixes, while vulnerability discovery and real-world exploitation continue to grow. Prioritization based on a single severity score often produces unstable rankings and ignores local context. Different scoring methods may also rank the same vulnerability differently. This paper presents an approach that combines multiple signals (vulnerability properties, current threat evidence, infrastructure context, and remediation feasibility) within attack-graph reasoning. Patch decisions are formulated as constrained multi-objective selection, with the aim of producing stable, explainable, and operationally usable patch plans under realistic limits such as maintenance windows, dependencies, and team capacity.**

***Keywords*—attack graphs, patch prioritization, vulnerability management, multi-objective optimization**

I. Introduction

In small networks, patch prioritization is not only about ranking vulnerabilities, but also about selecting a feasible set of fixes. In each cycle, only a limited number of patches can be deployed. Practical guidance therefore treats patching as continuous maintenance that requires planning, scheduling, and risk control [1]. In practice, prioritization based only on severity often fails to give stable and context-appropriate results, especially when scoring methods disagree [2] or when the real exposure depends on network reachability and service placement.

Input quality matters as well. Inconsistencies in public vulnerability reports can enter automated workflows and reduce decision quality [3]. This motivates a context-aware method that combines multiple signals and makes trade-offs explicit. The goal is not to replace severity metrics, but to complement them with information that small environments usually have (asset inventory, scanner output, basic topology) and to keep the decision process explainable.

II. Background: attack graphs and infrastructure-aware signals

Attack graphs model multi-step compromise by describing reachability, privilege changes, and dependencies between assets. They support system-level reasoning about which attack paths are feasible and how remediation changes those paths. Scalable attack-graph generation and abstraction are important because naive construction may be too large even for mid-sized infrastructures [4], [5].

Graph-based security metrics capture structural importance and possible risk propagation. Survey work reviews quantitative network security metrics and shows how graph-derived indicators can support prioritization [7]. In attack-graph settings, the chosen representation affects computed rankings. For this reason, representation comparison is important for stability [6]. Network-level aggregation of vulnerability scoring has also been studied to obtain richer system metrics for comparing remediation choices [8].

For small networks, attack graphs are most useful for *relative* decision support. They help identify which few fixes most reduce feasible attacker progress under a stated reachability model, rather than estimating an absolute *true* risk. This perspective also supports repeatable evaluation. Given the same inputs and assumptions, the method should produce comparable outputs, while uncertainty can be explored through controlled changes in reachability or missing attributes.

III. Approach overview

A. Signal taxonomy

The approach groups prioritization inputs into complementary signal families:

- **S1: Basic vulnerability properties** (severity and exploitability-related attributes).
- **S2: Current threat signals** (exploitation evidence, exploit availability, attacker targeting).
- **S3: Infrastructure context** (exposure, reachability, dependency structure, attack-path roles, risk aggregation).
- **S4: Remediation feasibility** (deployment time, operational risk, coordination cost, team load).

This structure reflects that severity alone cannot represent feasibility and local context, and that operational decisions require explicit trade-offs [1], [7]. It also supports incremental use. If threat signals are not available, the method can still operate using basic properties and infrastructure context, and later add threat signals without changing the overall model.

B. Infrastructure context via attack-graph metrics

Infrastructure context is represented through an attack graph and derived metrics, for example:

- **Path disruption:** how remediation removes or lengthens feasible attack paths, especially those leading to high-impact targets.
- **Bottlenecks and influence:** graph indicators that help identify critical points in attacker movement [7].
- **Representation sensitivity:** how stable the ranking remains across reasonable graph representations and modeling choices [6].

To stay practical for small networks, the method considers abstraction and reduction to keep graphs smaller and easier to interpret [5]. In practice, multiple graph signals can be computed (e.g., shortest-path disruption and influence-like measures) and treated as separate objectives, rather than forcing them into one scalar score.

C. Operational feasibility as first-class constraints

Small networks often face strict feasibility limits. These include short maintenance windows, limited staff, and high disruption cost. The approach treats feasibility as explicit constraints, aligned with patch management planning guidance [1]. This supports actionable outputs, which are selected patch sets that fit capacity and dependency constraints, rather than only a best-effort ranking.

Feasibility modeling can include: **(i)** a maximum number of changes per cycle, **(ii)** patch dependencies and prerequisites, **(iii)** estimated downtime, and **(iv)** expected coordination effort (e.g., vendor involvement). Even coarse estimates are useful when constraints are explicit, because they prevent recommendations that cannot be executed.

D. From ranking to selection: multi-objective formulation

Instead of producing only a total order, patch planning is formulated as selecting a subset of fixes that fits constraints. Objective components represent complementary goals such as threat pressure, criticality, risk spread, and team load. The output can be one recommended set under a chosen policy, or a small Pareto set of non-dominated alternatives with brief explanations.

A key point is that *best* is not unique. One patch set may maximize expected risk reduction, while another may achieve nearly the same reduction with much lower effort or lower downtime. Presenting a small set of strong options supports decision-making in small teams where deployment constraints dominate.

IV. Methodological components

A. Assumptions and input uncertainty

A core methodological goal is to separate uncertainty caused by incomplete inputs (inventory gaps, unclear exposure, missing versions) from uncertainty caused by modeling choices. Data-quality issues in vulnerability reports motivate explicit handling of inconsistencies and sensitivity analysis [3]. This makes evaluation more credible, because instability can be linked either to noisy inputs or to the method itself.

Operationally, this can be implemented as: **(i)** an explicit assumption set (e.g., default reachability when unknown), **(ii)** conservative vs. optimistic modes, and **(iii)** sensitivity tests where a small fraction of edges or affected versions is perturbed. The outputs should remain interpretable in both modes.

B. Signal-to-input mapping

Signals are mapped to data sources common in small networks: scanner findings and inventories (basic properties), threat feeds or advisories (threat signals), topology and reachability modeling (infrastructure context), and change-management constraints (feasibility). Attack graphs act as the integration layer where per-CVE attributes are combined with network semantics [4], [5].

When inputs are missing, the method should degrade gracefully. Missing threat evidence should not block computation. Instead, it should reduce confidence or shift emphasis to infrastructure context and feasibility. If asset criticality is unknown, impact can be approximated using simple tiers (e.g., *user endpoint* vs. *server* vs. *gateway*).

C. Metric interpretation and stability considerations

The approach uses established graph reasoning but treats stability as a primary requirement. Different attack-graph representations can produce different remediation rankings for the same candidates [6]. This motivates controlled experiments where reachability assumptions and representation choices are varied to quantify stability.

Stability is assessed at two levels. The first is stability of the *ordering*, meaning how much the ranking changes. The second is stability of the *selected set* under fixed capacity, meaning how many chosen patches change. In practice, set stability is often more important. Small teams patch a fixed number of items per cycle, so changes around the selection boundary matter most.

D. Feasibility-aware outputs

Feasibility is integrated so that recommended patch sets can be deployed. This aligns with patch management guidance [1] and supports short rationales for selected items, explaining why a patch is chosen in terms of context and cost trade-offs [7]. A practical output format is a small table with selected patch, affected asset, estimated effort, main reason (threat or context), and expected effect (path disruption or target protection).

E. Explainability and decision trace

To support adoption, each recommendation should include a compact decision trace, stating which signals had the strongest effect and which constraints were binding. For example, if a patch is not selected despite high severity, the trace can show that the asset is unreachable from the assumed entry points, or that the patch exceeds the maintenance-window budget. This improves trust and also helps improve inputs. If a decision depends on a reachability assumption, the operator can verify or update the network model.

V. Next steps and evaluation plan

A. Input model, graph construction, and scalability

Future work will implement a practical input model for small networks and choose an attack-graph construction strategy that is both tractable and interpretable. Established work on scalable generation and abstraction provides a basis for this design [4], [5]. Evaluation can also use vulnerable-network generators to test scalability under controlled variation [9].

The implementation stage will also define a baseline library of small-network topologies (e.g., flat LAN, segmented LAN with DMZ, remote access gateway) and a consistent mapping from scan outputs to graph nodes and edges. This is needed for repeatable experiments and for fair comparison of metric combinations.

B. Effectiveness, stability, and validity

Evaluation will compare the multi-metric selection against baselines such as severity-only and single-signal heuristics. Metrics will include: **(i)** reduction of reachable critical states and high-risk paths, **(ii)** risk reduction per unit effort, and **(iii)** stability under perturbations of inputs and scoring signals. Data-quality threats will be addressed using sensitivity analysis motivated by known inconsistencies in public vulnerability reports [3].

Threats to validity include imperfect mapping of vulnerabilities to affected services, incomplete reachability data, and simplified asset criticality. These will be mitigated by documenting assumptions, using multiple scenarios, and reporting results as ranges (best, typical, worst) rather than single-point claims.

C. Extensions: patch content

Beyond CVE-level signals, patch content and repair patterns can improve validation and remediation planning [10]. This can strengthen the link between vulnerability descriptions and actual remediation artifacts and can support better effort estimates. In small networks, even a coarse repair class (configuration change vs. package update vs. major version upgrade) can improve planning and reduce failed deployments.

D. Practical baselines and reporting

For comparability, baselines should include: **(i)** severity-only, **(ii)** severity plus reachability, and **(iii)** severity with feasibility constraint filtering. Reporting should include not only average improvements but also cases where the method matches baselines. This clarifies when multi-metric reasoning provides clear value and when available inputs are too weak to justify extra complexity.

VI. Conclusion

This paper described a multi-metric approach to patch prioritization for small networks that integrates heterogeneous signals into attack-graph context and frames patch planning as constrained selection. The approach emphasizes stability, interpretability, and deployability under realistic operational limits. The next step is implementation and repeatable evaluation focused on effectiveness and stability against one-dimensional baselines, with explicit treatment of input uncertainty and representation sensitivity.

A. Future work

Future work will follow the thesis goals by operationalizing four core signal groups and integrating them into a single decision process. First, the method will formalize *exploitation likelihood* as a threat-informed component that reflects whether a vulnerability is realistically likely to be exploited within the observed threat landscape. Second, it will incorporate *criticality* so that remediation decisions reflect the importance of affected assets and services in a small-network setting. Third, it will quantify *risk spread* through attack-graph structure, capturing how individual fixes reduce attacker reachability and limit multi-step compromise pathways. Fourth, it will model *team load* as an explicit operational dimension, representing effort, complexity, and deployment constraints that determine how many fixes can be executed in a given cycle.

These components will be combined through a multi-objective optimization formulation that produces either a single recommended patch set under a defined policy or a small set of Pareto-optimal alternatives. The evaluation will be designed to demonstrate that the resulting prioritization is both more effective and more stable than approaches based only on severity or on a single signal family. This includes controlled experiments that compare rankings and selected patch sets across different scenarios, and that measure stability under small changes in inputs and assumptions. The expected outcome is a practical methodology for small networks that supports repeatable comparison, produces actionable remediation plans, and provides concise explanations that link threat, context, and feasibility to the final patch selection.

Acknowledgment

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

References

[1] M. Souppaya and K. Scarfone, "Guide to Enterprise Patch Management Planning: Preventive Maintenance for Technology," NIST Special Publication 800-40 Rev. 4, 2021.

[2] V. Koscinski, M. Nelson, A. Okutan, R. Falso, and M. Mirakhorli, "Conflicting Scores, Confusing Signals: An Empirical Study of Vulnerability Scoring Systems," in *Proc. ACM CCS*, 2025.

[3] Y. Dong, W. Guo, Y. Chen, X. Xing, Y. Zhang, and G. Wang, "Towards the Detection of Inconsistencies in Public Security Vulnerability Reports," in *Proc. USENIX Security*, 2019.

[4] X. Ou, W. F. Boyer, and M. A. McQueen, "A Scalable Approach to Attack Graph Generation," in *Proc. ACM CCS*, 2006, pp. 336–345.

[5] S. Zhang, X. Ou, and J. Homer, "Effective Network Vulnerability Assessment through Model Abstraction," in *DIMVA*, LNCS, Springer, 2011, pp. 17–34.

[6] T. Gonda, T. Pascal, R. Puzis, G. Shani, and B. Shapira, "Analysis of Attack Graph Representations for Ranking Vulnerability Fixes," in *GCAI*, EPiC Series in Computing, 2018, pp. 215–228.

[7] A. Ramos *et al.*, "Model-Based Quantitative Network Security Metrics: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2704–2734, 2017.

[8] P. Cheng, L. Wang, S. Jajodia, and A. Singhal, "Aggregating CVSS Base Scores for Semantics-Rich Network Security Metrics," in *Proc. SRDS*, IEEE, 2012, pp. 31–40.

[9] A. Palma and S. Bonomi, "Behind the Scenes of Attack Graphs: Vulnerable Network Generator for In-depth Experimental Evaluation of Attack Graph Scalability," *Computers & Security*, vol. 157, p. 104576, 2025.

[10] X. Cao, J. Wang, and P. Wu, "Enhancing Vulnerability Repair through the Extraction and Matching of Repair Patterns," *Journal of Systems and Software*, vol. 230, p. 112528, 2025.

Advanced XR technologies, their implementation environments, and applications

[1]*Timotej SOBOTA (1st year),*
Supervisor: [2]*Branislav SOBOTA*

[1,2]Dept. of Computer and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]timotej.sobota@tuke.sk, [2]branislav.sobota@tuke.sk

***Abstract*—Extended Reality (XR) integrates Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) into a spectrum of interactive systems that can support training, visualization, remote collaboration, and simulation. This paper summarizes advanced XR technologies with an emphasis on how implementation environments (hardware platforms and software stacks) constrain real-world deployability. Key limitations are discussed through the lens of usage burden, including performance and robustness requirements (e.g. tracking reliability) and human-factor constraints such as discomfort and cybersickness. The paper further reviews multimodal interaction (hand tracking, voice, gaze, controllers, haptics) as a practical approach to increase interaction bandwidth and provide fallback behavior under sensing degradation. Finally, an iterative dissertation-oriented workflow is outlined, linking prototyping and task definition with user studies and analysis, with the goal of deriving experimentally grounded recommendations for building XR systems that are more robust, user-friendly, and deployable across application contexts.**

***Keywords*—virtual reality, augmented reality, mixed reality, extended reality, multimodality**

I. Introduction

The term Extended Reality (XR) is now widely recognized, although it was previously referred to as separate subsystems: Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR).

As Spadoni et. al [1] state, XR technologies have returned to captivate the industrial world, mainly due to the enabling technologies have improved to the point of pushing companies to re-invest in them.

XR technologies can be applied across a wide range of fields. From immersive learning in STEM [1] education, remote work to medical training and therapeutic intervention.

II. Advanced XR technologies

This renewed interest, as mentioned in Section I, in these technologies has also motivated my research in this technology together with my further academic focus. The motivation for reviewing advanced XR technologies is to determine whether these technologies are deployable beyond controlled demonstrations.

First, the practical value of XR in training, visualization, remote collaboration, and simulation depends on meeting strict system-level requirements (f.e. spatial tracking, low end-to-end latency etc.). If these requirements are not satisfied, the user experience degrades rapidly and can lead to reduced task performance or early termination of use. Second, XR adoption is strongly limited by *usage burden*, which includes physical comfort (e.g., fatigue), perceptual side effects (e.g., cybersickness), cognitive workload, interaction effort, and operational overhead such as calibration and setup. These factors directly constrain how long and in what contexts XR can be used.

[1]STEM - Science, Technology, Engineering, and Mathematics

III. Proposed Methodology

The proposed work will likely follow an iterative workflow with user-centered evaluation to study how multimodal interaction in XR can reduce usage burden while maintaining robustness and performance (Fig. 1).

Prototyping and task definition: The work should begin by implementing a baseline XR prototype together with multimodal variants. Next, a set of repeatable tasks must be defined, for example navigation, object selection/manipulation, and command or menu actions. Finally, the evaluation conditions will likely be specified to reflect real use, such as occlusion and increasing scene complexity.

User study: This step is expected to involve a user study for comparison of interaction conditions (basic versus multimodal). The order of conditions should be counterbalanced and breaks can be included to reduce fatigue and sickness effects. During the study, objective performance logs should be collected alongside subjective questionnaire responses.

Analyze results and iterate: The collected data are intended to be analyzed using task metrics (time, errors, interaction breakdowns, recovery time) and system performance indicators (FPS and frame stability). Subjective outcomes such as usability, workload, presence, and sickness can be assessed and complemented with qualitative feedback. The findings should then be used to refine the tasks, fusion/fallback logic, and UI design, after which the process returns to the prototyping phase for the next iteration.

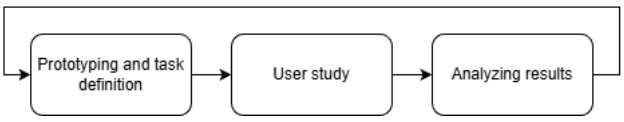


Fig. 1. Iterative workflow: prototyping and task definition, user study, and analysis feeding back into the next iteration.

By applying this methodology, it will be necessary to examine equivalent studies from around the world, as well as

the ones from our home workspace, the LIRKIS laboratory. For this reason, I will present the following chapters.

IV. State of the Art and Analysis

Advances in computing have accelerated the development of VR, AR, and MR, which are increasingly grouped under the umbrella term XR [2].

The categorization of XR systems includes a classification according to the degree of immersion, from simple XR systems that use only basic visual and audio effects, to highly immersive systems that include advanced technologies such as motion tracking, haptic suits, and biofeedback. This categorization helps in selecting the right technology for a specific purpose.

A. Definition of subsystems

VR is commonly defined as an immersive, three dimensional computer-generated virtual environment that replaces the user's visual surroundings and supports real-time interaction (Fig. 2), as stated in [3]. The user can be fully immersed in this virtual environment either by using a VR headset or enter a virtual cave (Fig. 5).

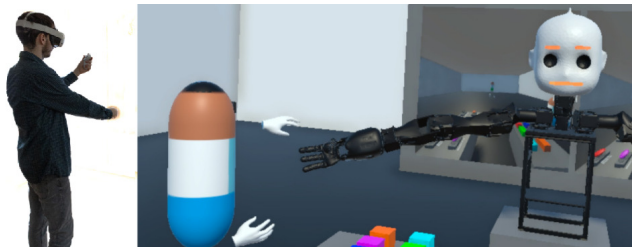


Fig. 2. A third person showcase of VR, where the user wears a VR headset and is controlling his avatar.[4]

In contrast, *AR* overlays digital content onto the user's view of the real world (f.e. by using a phone as shown in Fig. 3), enabling simultaneous perception of physical and virtual elements and interaction with both [5].

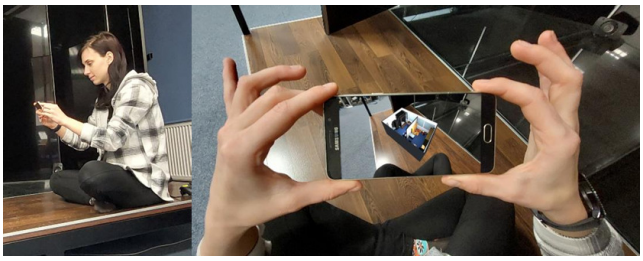


Fig. 3. Showcase of AR, performed in LIRKIS lab.[6]

MR is often positioned between VR and AR. As Sobota et. al [3] state: it can not only place virtual elements into the physical environment, but it also enables more consistent gestural interaction between virtual objects and physical space. Fig. 4, showcases the use of Microsoft Hololens, a head-worn, optically see-through display [7], in LIRKIS lab.

While these categories are conceptually distinct, practical XR systems differ primarily in how they:

1) acquire information about the user and environment (e.g., head pose, hand pose, depth/scene understanding),
2) render content with correct spatial alignment and low latency,

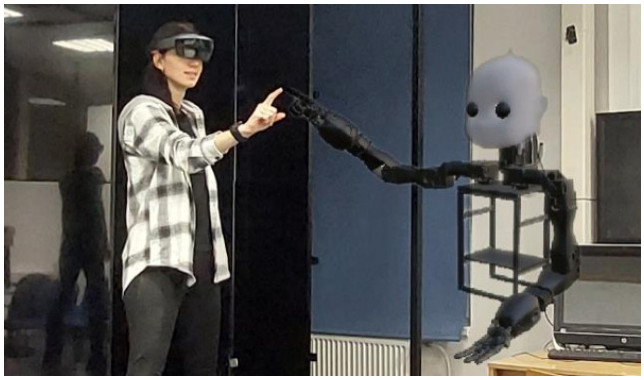


Fig. 4. Showcase of MR, performed in LIRKIS lab using HoloLens.[6]

3) support interaction through controllers or natural input.

VR typically emphasizes immersion and presence, but it can increase physical and cognitive load due to sensory conflicts and limited awareness of the real environment, which can lead to cybersickness. AR and MR preserve situational awareness, yet they place stronger requirements on robust tracking, occlusion handling, and stable registration under variable lighting and dynamic scenes.

Collaborative XR systems allow users from different geographic locations to collaborate in a single virtual environment. These systems can be used for team-building activities, joint training, or even for collaborative work on engineering and design projects. Collaborative VR and XR require robust networks and platforms to ensure data synchronization and interactions between participants, which includes audiovisual streaming and real-time data sharing. *Locally shared XR systems* enable user interaction within a limited geographical space, such as rooms or buildings, while globally shared systems extend collaboration opportunities to an international level, supporting users located around the world. This ability to connect people across different locations opens up new possibilities for virtual offices, educational programs, and simulated training environments where participants can interact as if they were in the same physical space.

B. Key limitations tied to UX

As mentioned above, major user experience (UX) limitation of immersive VR is cybersickness, commonly defined as a constellation of discomfort symptoms (e.g., nausea, disorientation, eye strain) induced by VR exposure. As Weech et. al [8] state, cybersickness is strongly linked to sensory integration and sensory mismatch mechanisms, particularly conflicts between visual motion cues and the users body position, and it can directly impair task performance or lead users to terminate sessions early. Importantly, the literature also indicates that cybersickness and the desirable experience of presence are often negatively related, suggesting that discomfort can distract users and reduce involvement in the virtual environment.

As Lawson et. al [9] state, cybersickness is influenced by three broad groups of factors: (1) stimulus/display characteristics, (2) user behavior in the virtual environment, and (3) individual/user characteristics.

1) *Stimulus and display characteristics*
   The report explicitly highlights that the "sickening challenge" depends on characteristics of the stimulus, including:

- the amount of real or visual motion driving unusual sensorimotor integration,
- duration of exposure,
- display refresh rate,
- field of view (FOV).

2) User behavior during XR use
Cybersickness also depends on what the user does while in the synthetic environment, especially:
   - the number and type of head/body movements required to complete tasks or perform virtual locomotion.

3) Individual differences (user characteristics)
There is substantial variability in susceptibility across users. Reported contributors/predictors include:
   - inherent susceptibility differences across users,
   - experience or familiarity can change tolerance,
   - illness or reduced health state (e.g., fatigue, sleep loss, head colds, hangover, stress),
   - device fit or adjustment issues,
   - anxiety/personality factors (state anxiety may contribute to nausea and discomfort),
   - history of prior symptoms (past symptom severity as a practical predictor of future severity).

As Korečko et. al [10] emphasize, questionnaires are an important part of subjective UX evaluation methods, capturing the user's personal perspective on a system. They also note that many UX questionnaires exist and some have become standards. In connection with previous work based in the LIRKIS lab, an evaluation of the mentioned UX testing will also be carried out using questionnaires. It is considered good practice to use SUS and Nasa TLX questionnaires, which have a broad applicability and can also be used for non-immersive XR. Furthermore, VRSQ and IPQ should be used for immersive XR. In summary, it is assumed that the following types of questionnaires will be used in the questionnaire form:

- System Usability Scale (SUS) (usability)
- NASA Task Load Index (NASA-TLX) (workload)
- Virtual Reality Sickness Questionnaire (VRSQ) (negative aspects / sickness)
- Igroup Presence Questionnaire (IPQ) (presence)

The use of technology from a technical and software perspective will be described in the following subsections (Sec. IV-C and Sec. IV-D). It is assumed that the use of multimodality will also make it possible to obtain other accurate data for evaluating stress in the XR environment, such as subjective questionnaires.

C. Implementation environments (HW/SW)

XR hardware is commonly characterized and compared using device form factor and sensing/display specs. XR devices can be categorized by typology (head-mounted vs smartphone/tablet-based).

The head-mounted display (HMD) is the main and probably the most commonly used piece of hardware (HW) to enable high imersion in XR for the user as mentioned in [11]. Immersion and perceived quality depend on multiple HW-linked parameters. Casini's review [12] notes that immersivity depends on characteristics such as FoV, display quality (e.g., pixel density), refresh rate, allowed motion (DoF), tracking accuracy, controllers, and audio.

A representative example of a large-scale XR implementation environment is the LIRKIS CAVE (Cave Automatic Virtual Environment) developed at the Technical University of Košice [13]. Unlike head-mounted displays, the LIRKIS CAVE uses a multi-display room-like setup in which the display surfaces are covered by stereoscopic LCD panels and user motion is tracked by an optical tracking system (Fig. 5). Real-time rendering is distributed across a small

Fig. 5. Showcase of the LIRKIS cave located in LIRKIS lab.[6]

compute cluster. From an implementation perspective, LIRKIS CAVE highlights how XR performance and robustness depend not only on rendering fidelity, but also on synchronization across rendering nodes, tracking quality, and interaction-device integration within a heterogeneous HW/SW stack.

When it comes to XR software (SW), the content creation and interaction typically relies on specialized real-time engines rather than traditional 3D/BIM[2] tools. Casini [12] explains that dynamic XR requires continuous content generation using real-time 3D rendering tools, and that this performance level "cannot be achieved by conventional 3D and BIM rendering software," motivating the use of optimized graphics engines integrated with interactions via libraries and headset platform Software Development Kits (SDKs). For example: AR SDKs such as Vuforia or OpenXR , and identifies common engines for XR visualization/interaction (e.g., Unity, Unreal Engine, etc.).

Middleware and system architecture are key "software glue" for XR in heterogeneous environments. Simões and others [14] argue that industrial XR adoption is hindered by low digitalization, infrastructure complexity, and repeated calibration needs, and they propose a middleware architecture that enables communication and interaction across technologies without manual configuration or calibration, including mechanisms for automated calibration and data format transformation between devices.

Edge/Fog/Cloud placement affects XR feasibility when latency and data volume are constraints. [14] also states that cloud-only IT architectures may fail strict industrial requirements (e.g., large data volumes with low latency), motivating migration toward Fog/Edge infrastructures and software architectures that handle real-time processing and streaming close to devices.

D. Multimodality

Multimodality has become a central design principle in contemporary XR systems because no single interaction channel remains reliable and efficient across all users, tasks, and environments. The authors of a scoping review on this topic

[2]BIM - Building Information Modeling

[15] note that current XR systems already provide highly realistic visual and auditory experiences, and that low-cost HMD-based systems are beginning to stimulate additional senses. However, beyond vision and audio, the most common deployed method is still relatively simple vibrotactile haptics via hand-held controllers, leaving many human perceptual capabilities underused.

From both an Human-Computer Interaction (HCI) and systems perspective, multimodality in XR spans multiple input and output modalities:

1) *Hand tracking (gestures)* - natural interaction for pointing, grabbing, pitching etc. It can also reduce dependence on controllers for simple tasks.
2) *Voice and speech commands* - fast command invocation ("open menu", "reset view"), hands-free control, and accessibility benefits. Particularly useful when hands are busy (manipulation tasks) or when tracking quality degrades.
3) *Gaze and eye tracking* - Rapid target acquisition (select what you look at). Can reduce controller/hand precision requirements by using gaze for "pointing" and another modality for "confirm".
4) *Controller-based input (buttons, joysticks, tracked controllers)* - High reliability and precision for selection, locomotion, and object manipulation.
5) *Haptics* - Confirms contact events (clicks, hits, collisions), improves realism in manipulation, and can reduce visual dependence for feedback.

In a 2025 study on multimodality in XR [16], these modalities are evaluated empirically using a hybrid methodology (user testing with both quantitative and qualitative measures) and standardized questionnaires including IPQ and UEQ-S, demonstrating that modality choices can measurably affect both pragmatic (usability/efficiency) and hedonic (enjoyment/appeal) aspects of user experience.

V. Conclusion

This paper summarized advanced XR technologies with emphasis on the practical constraints that determine whether XR systems are deployable outside controlled demonstrations. Across application domains, real-world adoption is shaped by the combined impact of implementation environments and user-centered limitations captured by the notion of *usage burden*. In particular, stable performance and robust tracking/registration are not only engineering goals but prerequisites for reliable interaction and acceptable user experience.

The reviewed literature also indicates that multimodal interaction is a promising direction for improving both usability and robustness. By combining modalities such as hand tracking, gaze, controllers and haptics, XR systems it may increase interaction bandwidth and maintain functionality when individual sensing channels degrade. However, multimodality introduces additional design challenges related to fusion logic, ambiguity resolution, and potential cognitive overload, which motivates systematic evaluation.

Based on these findings, the proposed dissertation will focus on iterative prototyping and user-centered evaluation of multimodal XR workflows under realistic stressors (e.g., occlusion, scene complexity). The expected outcome is a set of experimentally grounded design recommendations and implementation strategies that reduce usage burden while preserving robustness, supporting more dependable XR deployment in practical applications.

Acknowledgment

This work has been supported by the APVV grant no. APVV-21-0105 "Trustworthy human–robot and therapist patient interaction in virtual reality".

References

[1] E. Spadoni, M. Bordegoni, M. Carulli, and F. Ferrise, "Extended reality in industry: past, present and future perspectives," *Proceedings of the Design Society*, vol. 3, pp. 1845–1854, 2023. [Online]. Available: https://doi.org/10.1017/pds.2023.185

[2] B. Sobota, Š. Korečko, I. Farkaš, J. Kolesár, and T. Sobota, "Extended reality based fire safety training system for humanoid robot knowledge base," in *2025 International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 2025, pp. 754–759. [Online]. Available: https://doi.org/10.1109/ICETA67772.2025.11280134

[3] B. Sobota, M. Hudák, and E. Pietriková, "Collaborative xr systems and computer games development," in *Computer Game Development*. IntechOpen, 2022. [Online]. Available: https://doi.org/10.5772/intechopen.105555

[4] T. Sobota, "Using a metaverse platform to present trustworthy human–robot communication," Master's thesis, Technical University of Košice, Faculty of Electrical Engineering and Informatics, Košice, Slovakia, 2025, supervisor: Assoc. Prof. Branislav Sobota, PhD. (in slovak).

[5] R. T. Azuma, "A survey of augmented reality," *Presence: teleoperators & virtual environments*, vol. 6, no. 4, pp. 355–385, 1997. [Online]. Available: https://doi.org/10.1162/pres.1997.6.4.355

[6] M. Mattová, "Advanced virtual reality technologies," Doctoral dissertation, Technical University of Košice, Faculty of Electrical Engineering and Informatics, Košice, Slovakia, 2024, study programme: Informatics. Department: Department of Computers and Informatics (KPI). Supervisor: Assoc. Prof. Branislav Sobota, PhD. (in slovak).

[7] C. Gsaxner, J. Li, A. Pepe, Y. Jin, J. Kleesiek, D. Schmalstieg, and J. Egger, "The hololens in medicine: a systematic review and taxonomy," *Medical Image Analysis*, vol. 85, p. 102757, 2023. [Online]. Available: https://doi.org/10.1016/j.media.2023.102757

[8] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: a review," *Frontiers in psychology*, vol. 10, p. 158, 2019. [Online]. Available: https://doi.org/10.3389/fpsyg.2019.00158

[9] B. D. Lawson, P. Proietti, S. Peder, R. Timothy, K. Ramy, and B. Marten, "Factors impacting cybersickness," 2021. [Online]. Available: https://lib.iitta.gov.ua/id/eprint/729263/1/TR-HFM-MSG-323-ALL-73-120.pdf

[10] Š. Korečko, J. Mathes, and B. Sobota, "The hololens in medicine: a systematic review and taxonomy," *Proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice*, vol. 16, pp. 582–587, 2025, (in slovak). [Online]. Available: https://eei.fei.tuke.sk/wp-content/uploads/2025/07/EEI_16.pdf

[11] A. Alnagrat, R. C. Ismail, S. Z. S. Idrus, and R. M. A. Alfaqi, "A review of extended reality (xr) technologies in the future of human education: Current trend and future opportunity," *Journal of Human Centered Technology*, vol. 1, no. 2, pp. 81–96, 2022. [Online]. Available: https://doi.org/10.11113/humentech.v1n2.27

[12] M. Casini, "Extended reality for smart building operation and maintenance: A review," *Energies*, vol. 15, no. 10, p. 3785, 2022. [Online]. Available: https://doi.org/10.3390/en15103785

[13] Š. Korečko, M. Hudák, and B. Sobota, "Lirkis cave: architecture, performance and applications," *Acta Polytechnica Hungarica*, vol. 16, no. 2, pp. 199–218, 2019. [Online]. Available: https://doi.org/10.12700/APH.16.2.2019.2.12

[14] B. Simões, R. De Amicis, I. Barandiaran, and J. Posada, "X-reality system architecture for industry 4.0 processes," *Multimodal technologies and interaction*, vol. 2, no. 4, p. 72, 2018. [Online]. Available: https://doi.org/10.3390/mti2040072

[15] I. Rakkolainen, A. Farooq, J. Kangas, J. Hakulinen, J. Rantala, M. Turunen, and R. Raisamo, "Technologies for multimodal interaction in extended reality—a scoping review," *Multimodal Technologies and Interaction*, vol. 5, no. 12, p. 81, 2021. [Online]. Available: https://doi.org/10.3390/mti5120081

[16] T. Kojic, M. Vergari, M. Warsinke, D. Ali, S. Möller, and J.-N. Voigt-Antons, "Multimodal user experience in extended reality: Exploring hand tracking, voice, and passthrough interactions," in *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems*, 2025, pp. 8–14. [Online]. Available: https://doi.org/10.1145/3712677.3720459

Bayesian Multi-Sensor AMP-vSLAM for Robust Industrial AR

[1]*Ivana NOVÁKOVÁ (4th year)*
Supervisor: [2]Miroslav MICHALKO

[1,2]Dept. of Computers and Informatics, FEEI, Technical University of Košice, Slovak Republic

[1]ivana.novakova2@tuke.sk, [2]miroslav.michalko@tuke.sk

***Abstract*— Based on the previously proposed prototype of adaptive multiphysics vSLAM (AMP-vSLAM), this paper further describes a Bayesian multisensor extension that explicitly models the reliability of individual sensor modalities in a dynamic industrial environment, supporting the successful implementation of augmented reality technology in simulation scenarios.**

***Keywords*—AMP-vSLAM, AR, Bayesian Multi-sensor, Multisensor fusion, vSLAM**

I. INTRODUCTION

Bayesian multi-sensor vSLAM is an approach to localization and mapping that combines information from multiple heterogeneous sensors [1] within an integrated probability framework. In SLAM, effectively fusing data from multiple sensors, such as vision, LiDAR, and IMU, is key to improving robustness and accuracy [6].

Instead of the traditional SLAM approach with fixed, static weights for individual sensors [3], their contributions to the state estimate are modeled as random variables that reflect the current reliability of their measurements. The problem is that in challenging environments such as manufacturing halls, classical visual-inertial SLAM algorithms [2] suffer significant performance degradation under intermittent lighting, mechanical vibrations, thermal drift, and dust, leading to drift, tracking failures, and unstable holograms.

This paper focuses on an illustrative Bayesian multi-sensor AMP vSLAM formulation and a simulated experiment that highlights how reliability-aware fusion can leverage additional industrial sensors to improve robustness.

II. RELATED WORK

Industrial environments are not scalable, as they include assembly-line manufacturing plants, blast-furnace steel mills, and underground mines. Each environment has specific conditions that significantly affect sensor reliability. Multi-sensor fusion and the development of a robust algorithm that can respond to rapidly changing conditions constitute a sound framework[4] for the successful implementation of AR in industry.

In our previous work, we presented the first functional prototype of AMP-vSLAM – an adaptive multiphysics visual SLAM developed for augmented reality applications. In this article, we take a closer look at a specific layer of multisensor fusion in the AMP-vSLAM we present. It includes a self-learning module that continuously adjusts each sensor's weight based on its current data quality. Bayesian reliability analysis continuously evaluates visual, inertial, thermal, and acoustic streams; noisy sensors are automatically filtered out, while more stable ones have greater impact. All measurements and contextual information are archived; machine learning models use this history to identify the most effective sensor weights under similar conditions and immediately apply the predicted coefficients to the next round of data fusion. Our initial testing showed positive results: 48% trajectory error and a feature retention rate up to 2.1 times that of ORB-SLAM3, while maintaining real-time overlay on the MS HoloLens 2 head-mounted display[4]. By combining visual, IMU, [5] and other sensors via Bayesian weighting and temperature-sensitive calibration, the system maintains map integrity even in the face of significant lighting changes, mechanical shocks, and thermal drift, [3] which are common in industrial environments.

III. EXPERIMENTAL RESULTS

The proposed AMP-vSLAM algorithm, despite positive results from our initial implementation, relied on several manually tuned, environment-specific fusion parameters, which limited portability across different lines or machines. To further eliminate errors, we chose a Bayesian multisensor vSLAM framework. This approach defines a common probabilistic model for the trajectory, environment map, sensor reliability, and all acquired measurements.

$$p(x_{0:T}, m, r_{0:T} \mid z_{1:T}),$$

$$z_t^s = h^s(x_t, m) + \epsilon_t^s, \epsilon_t^s \sim \mathcal{N}(0, \Sigma^s / r_t^s),$$

It enables adaptive handling of rapidly changing conditions in more demanding industrial environments. At each processing step, the system's prior beliefs about location, map, and sensor reliability are combined with new measurements. The probability distribution (the posterior) is then repeatedly updated, directly determining how strongly each modality contributes to the final state estimate.

For our experimental analysis, we model all sensor noise as Gaussian to cast the sensor fusion problem within a closed probabilistic framework, which allows us to use standard

Bayesian estimators, such as extended Kalman filters or factor-graph optimization. Specifically, the measurement errors of the camera and IMU under nominal conditions are described by

$$\epsilon_t^{cam} \sim \mathcal{N}(0, \sigma_{cam}^2),\ \sigma_{cam} = 0{,}02\ m;$$

$$\epsilon_t^{imu} \sim \mathcal{N}(0, \sigma_{imu}^2),\ \sigma_{imu} = 0{,}05\ m$$

The simulation environment is designed to reflect a typical manufacturing floor with dynamic lighting conditions, machine-induced vibrations, thermal fluctuations, and dust. The selected sensors include a monocular RGB camera, an IMU (triaxial accelerometer and gyroscope), a 2D LiDAR scanner, UWB localization beacons, vibration sensors, and thermal probes mounted on a robotic platform. The sensor outputs are generated by sampling from zero-mean Gaussian noise with standard deviations as shown in Table 1: 0.02 m for camera position measurements at nominal illumination (increasing to 0.10 m after visual degradation), 0.05 m for IMU-integrated position, 0.03 m for LiDAR range, and 0.08 m for UWB absolute position. These noise parameters are chosen to be consistent with the centimeter-level localization and depth-sensing accuracy reported for HoloLens 2 mixed-reality devices, while remaining general enough to be representative of other industrial AR platforms. The vibration and thermal channels provide auxiliary measurements with normalized noise levels of 0.06 m and 0.07 m, respectively, that primarily modulate the inferred sensor confidence rather than directly constrain position. All modalities are synchronized at 50 Hz, except for UWB, which is synchronized at 10 Hz, and changes in confidence or bias are triggered by scripted environmental events (conveyor start, heating, welding arc, and UWB activation). These noise parameters are chosen to be consistent with the centimeter-level localization and depth-sensing accuracy reported for HoloLens 2 mixed-reality devices, while remaining general enough to be representative of other industrial AR platforms.

Table 1 summarizes the resulting RMS localization errors for a conventional fixed-weight fusion scheme and the proposed AMP-vSLAM under the simulated disturbance scenario.

Simulation phase	**σ_{cam}[m]**	**σ_{imu}[m]**	**Weights (cam / IMU)**	**RMS error – fixed fusion [cm]**	**RMS error – AMP-vSLAM [cm]**
Before disturbance (t < 10 s)	0.02	0.05	0.7 / 0.3	2.05	2.05
After disturbance – fixed	0.10	0.05	0.7 / 0.3	7.16	–
After disturbance – adaptive*	0.10 → 0.224	0.05	0.047 / 0.953	–	4.88

Table 1 quantifies the impact of visual degradation on the RMS localization accuracy for conventional fixed-weight fusion and the proposed AMP-vSLAM framework. While under nominal conditions, both approaches provide comparable performance, achieving an RMS error of approximately 2.05 cm, when visual conditions deteriorate, the variance of the camera measurements increases significantly. In the fixed-weight scheme typical of vSLAM, the camera weight remains constant (w_cam = 0.7), leading the estimator to over-rely on the biased visual measurements and, as a result, increasing the RMS error to 7.16 cm.

The presented AMP-vSLAM, which includes Bayesian confidence estimation, adaptively scales the effective measurement covariance. After the conditions are disturbed, the camera's reliability decreases, leading to a standard deviation and subsequent weight reduction of approximately 4.7%. Kalman updating, therefore, shifts the credence towards inertial measurements, which remain statistically consistent. This adaptive covariance modulation reduces the RMS error to 4.88 cm, corresponding to approximately a 32% improvement over the fixed-weight baseline. Although AMP-vSLAM does not fully preserve sub-3 cm accuracy under the severe visual degradation required for hologram anchoring in a real-world environment [7], it significantly mitigates increases in errors and shortens the duration of perceptible hologram misalignment.

The assumed noise characteristics of the main sensors are summarized by the following standard deviations: the camera measurement noise is set to $\sigma_{\text{cam}} = 0.02$ m under nominal conditions and increases to 0.10 m after visual degradation; the inertial measurement unit exhibits a standard deviation of $\sigma_{\text{imu}} = 0.05$ m; the LiDAR sensor is characterized by $\sigma_{\text{lidar}} = 0.03$ m; and the UWB localization system by $\sigma_{\text{uwb}} = 0.08$ m. Vibration and thermal channels do not estimate position directly, but provide auxiliary measurements with typical noise levels $\sigma_{\text{vib}} = 0.06$ m (after normalization) and $\sigma_{\text{therm}} = 0.07$ m, which are used to infer sensor reliability.

Table 2 sensors.

Sensor	**σ [m]**	**description**
Camera	0.02 → 0.10	Degrades after lighting disturbance
IMU	0.05	Stable, accumulates drift
LiDAR	0.03	Robust to lighting
UWB	0.08	Absolute but sparse
Vibration	0.06	Correlated with mechanical activity
Thermal	0.07	Indicates environmental stress

To illustrate the effect of Bayesian reliability weighting, we simulate a 20 s trajectory and subject the system to gradually worsening conditions

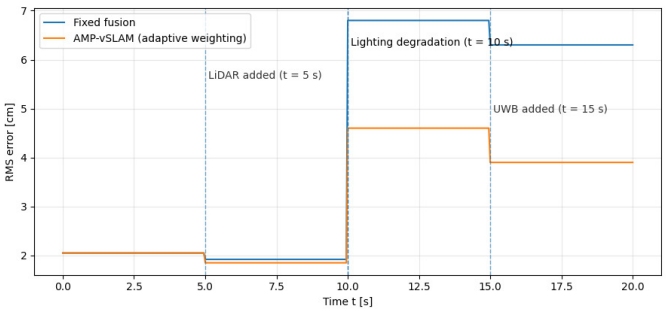
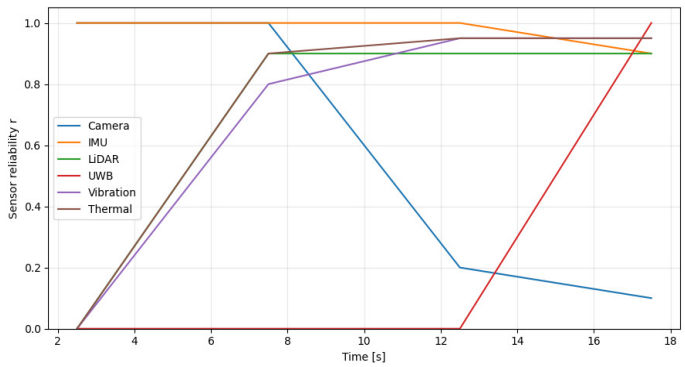


Figure 1RMS localization error in time-fixed fusion vs AMP-vSLAM

Table 3 shows the assumed reliability coefficients $r_t^s \in [0,1]$ for each modality in four time intervals. In the first 5 s, all sensors operate nominally. Between 5–10 s, conveyor-induced vibrations and moderate heating appear, slightly lowering the confidence in the IMU and increasing the relevance of vibration and thermal channels. Between 10–15 s, a strong welding arc causes severe glare and smoke, which drastically reduces camera reliability while other modalities remain high. Finally, between 15–20 s, an external UWB system becomes available and provides an additional absolute position constraint.

Table 3

Time [s]	Camera	IMU	LiDAR	UWB	Vibration	Thermal
0–5	1.0	1.0	–	–	–	–
5–10	1.0	1.0	0.9	–	0.8	0.9
10–15	0.2	1.0	0.9	–	0.95	0.95
15–20	0.1	0.9	0.9	1.0	0.95	0.95



In a conventional fixed-weight fusion scheme, additional sensors such as vibration and thermal measurements have no direct influence on the state estimate, even though they become highly informative in harsh industrial conditions. Our extended AMP-vSLAM framework incorporates these auxiliary modalities via Bayesian confidence estimation: elevated vibration or temperature levels automatically decrease the inferred reliability of the camera and IMU, while LiDAR and UWB are promoted when available. As a result, the system reduces the impact of corrupted visual measurements and enhances robustness to environmental disturbances, which is consistent with the improvements observed in our real-world experiments.

IV. Conclusion

This paper presented a Bayesian multi-sensor AMP-vSLAM framework that integrates camera, IMU, LiDAR, UWB, vibration, and thermal data through reliability-aware fusion to improve the robustness of industrial augmented reality localization and mapping. A simulated scenario showed that latent reliability variables enable the system to automatically downweight degraded visual measurements and leverage auxiliary sensors, reducing RMS localization error compared to fixed-weight fusion and stabilizing hologram placement under harsh factory conditions.

Building on this foundation, future work will incorporate a sophisticated meta-learning module that observes sequences of operating conditions and learns to predict reliability priors and adaptation rates for each modality across different machines and production lines. Such a meta-learner can optimize fusion hyperparameters for fast on-site adaptation, reuse experience from past AR deployments, and gradually improve robustness as more real-world AMP-vSLAM data are collected in industrial environments.

Acknowledgment

This publication was realized with the support of the Operational Programme Integrated Infrastructure, within the project Intelligent systems for UAV real-time operation and data processing (ITMS2014+: 313011V422), which was co-financed by the European Regional Development Fund.

References

[1] J. Zhu, H. Li and T. Zhang, "Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey," in *Tsinghua Science and Technology*, vol. 29, no. 2, pp. 415-429, April 2024, doi: 10.26599/TST.2023.9010010.

[2] Yu Xia, Hongwei Wu, Liucun Zhu, Weiwei Qi, Shushu Zhang, Junwu Zhu,A multi-sensor fusion framework with tight coupling for precisepositioning and optimization,Signal Processing, Volume 217,2024,109343,ISSN 0165-1684,https://doi.org/10.1016/j.sigpro.2023.109343.

[3] Yuan Zhuang, Xiao Sun, You Li, Jianzhu Huai, Luchi Hua, Xiansheng Yang, Xiaoxiang Cao, Peng Zhang, Yue Cao, Longning Qi, Jun Yang, Nashwa El-Bendary, Naser El-Sheimy, John Thompson, Ruizhi Chen,Multi-sensor integrated navigation/positioning systems using data fusion: From analytics-based to learning-based approaches,Information Fusion,Volume 95,2023, Pages 62-90, SSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.01.025.

[4] NOVÁKOVÁ, Ivana. 2025. 25th Scientific Conference of Young Researchers (SCYR): Proceedings from Conference. Košice: Technical University of Košice, 2025, s. 187–189. ISBN 978-80-553-4826-1.

[5] Leila Farahzadi, Ibrahim Odeh, Mahdi Kioumarsi, Behrouz Shafei, Automated image-based condition assessment of the built environment: A state-of-the-art investigation of damage characteristics and detection requirements,Results in Engineering,Volume 26,2025,104978,ISSN 2590-1230,https://doi.org/10.1016/j.rineng.2025.104978.

[6] Wang, Shaohui & Li, Fei. (2025). Design and Implementation of Efficient SLAM System Based on Multi-Sensor Fusion. Journal of Computer Science and Artificial Intelligence. 2. 22-27. 10.54097/hbb73g96.

[7] Uzam Haider, "Applications of Augmented Reality in Industrial Manufacturing in the Era of Industry 5.0", *International Journal of Engineering and Applied Physics*, vol. 5, no. 1, pp. 1127–1135, Jan. 2025.

Adaptive Sensor Fusion for Semantic Corner Case Detection Using Agentic AI

[1]*Matúš Dopiriak (4^{th} year),*
Supervisor: [2]*Juraj Gazda*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]matus.dopiriak@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—Current autonomous vehicle (AV) perception models frequently underperform in rare, safety-critical corner cases, posing heightened risk to other road users. We propose an agentic edge pipeline that adaptively fuses RGB and LiDAR for semantic corner-case reasoning. At the edge, a pretrained large visual language model (LVLM) converts RGB batches into RDF triples with confidence scores. If the adjusted confidence drops below a threshold, the server requests LiDAR bird's-eye-view (BEV) maps and re-queries the LVLM with an augmented RGB+LiDAR prompt, limiting LiDAR transmission and compute to uncertain cases. Our framework runs zero-shot without any retraining. On 3CSim and KITTI, our system improves corner-case semantic F_1 (up to +24% in low visibility) while reducing average data transmission by 35%, producing ontology-aligned triples for V2X sharing.**

***Keywords*—Adaptive sensor fusion, agentic AI, corner case detection, semantic communication.**

I. Introduction

Corner cases are rare and safety-critical scenarios that often fall outside standard training distributions, making autonomous vehicle (AV) perception brittle [1], [2]. LVLMs can improve corner-case understanding, but their computational cost motivates an edge-assisted deployment. This shifts the bottleneck to multimodal transmission and preprocessing: a 1920×1080 RGB frame is ∼926 KB (∼150 ms processing), while LiDAR adds ∼0.6–1.1 MB and ∼200 ms overhead. Sending both modalities can exceed ∼1.5 MB and ∼350 ms per instance, whereas RGB-only is faster but fails under low visibility where LiDAR is beneficial [3], [4]. We therefore target efficient multimodal pipelines that retain robustness while meeting real-time bandwidth and latency constraints.

We propose a lightweight fusion-based corner-case reasoning system for autonomous driving. It uses an agentic AI two-stage pipeline that performs fast RGB-only inference and dynamically invokes LiDAR when semantic confidence drops, balancing latency with robustness. To ensure semantic interoperability, it relies on a shared closed-world ontology and encodes scene interpretations as ontology-aligned RDF triples (e.g., `(ex:Pedestrian1, avcco:violates, avcco:RedLight)`), enabling zero-shot consistency checking and actionable decisions as in Fig. 1.

In summary, our contributions are:

1) An *agentic*, ontology-guided framework for zero-shot semantic corner-case detection in AVs with RDF triple outputs.
2) A confidence-aware fusion pipeline that selectively invokes LiDAR to balance robustness, latency, and bandwidth.
3) An evaluation on synthetic (3CSim) and real (KITTI) datasets showing improved low-visibility F_1 and reduced transmission.

II. Related Work

AI agents are promising for corner-case reasoning [5], but prior work largely assumes RGB-only sensing and ideal conditions [6], [7], limiting evaluation under adverse weather or sensor degradation. Our framework remedies this with adaptive RGB+LiDAR fusion and systematic testing across diverse corner cases and weather, revealing which modalities matter per scenario.

While sensor-fusion datasets have been explored for corner-case detection [8], they typically do not optimize sensor usage. It prioritizes informative sensors via confidence-driven fusion, reducing transmission and compute.

Real-world corner-case datasets (e.g., CODA) [9] often lack matched scenarios across environment conditions; we use 3CSim [1] to enable controlled replication under different weathers for sensor-necessity studies.

Finally, we improve agent accuracy without retraining by grounding reasoning in external knowledge (RAG [10]) and a domain ontology, represented as RDF triples for scenario-specific context [11].

III. Proposed Approach

A. Problem Definition

AVs face rare, ambiguous corner cases (e.g., occluded pedestrians or low-light debris) that can break standard perception [1], [2]. We model a coupled pipeline: sensors $S = \{s_1, \ldots, s_n\}$ (RGB, LiDAR) capture complementary views, a fusion function f_{fuse} merges streams (without per-frame calibration), and a reasoning function f_{reason} uses ontology O to produce RDF triples for V2X sharing.

This pipeline must satisfy stringent real-time and resource constraints: end-to-end latency must remain under 200ms per frame, per-frame bandwidth should not exceed 2MB to avoid network congestion, fusion must be robust without costly offline calibration, and all AVs must interpret corner cases identically under the same closed-world ontology O. Formally, we model an AV as the tuple:

$$A = \big(S,\ f_{\text{fuse}},\ O,\ f_{\text{reason}}\big), \tag{1}$$

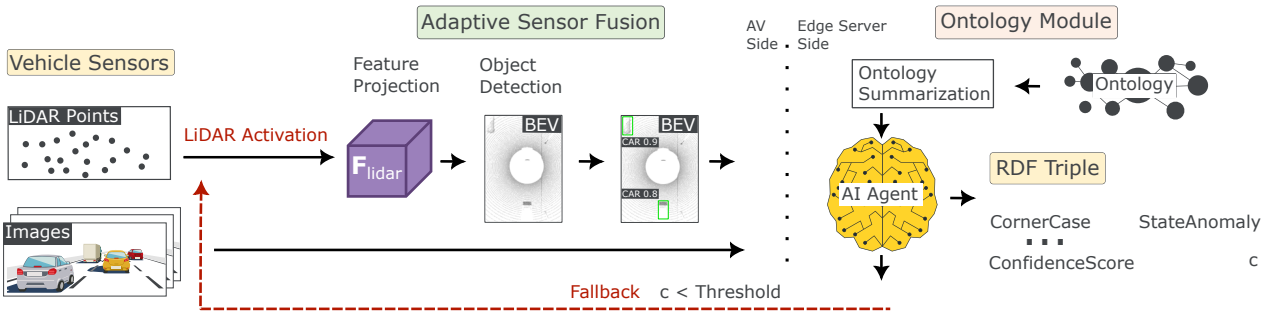


Fig. 1. The system architecture. RGB images are first sent from the AV to an edge-hosted AI agent for ontology-guided RDF triple generation. If semantic confidence falls below 85%, the AV triggers a fallback, sending additional LiDAR BEV inputs for refined multimodal reasoning.

where S is the set of multimodal sensors, $f_{\text{fuse}} : S \to \mathcal{X}$ maps raw sensor data to a fused latent representation $\mathcal{X}$, O is a shared corner-case ontology encoding classes, properties, and relations, and $f_{\text{reason}} : (\mathcal{X}, O) \to T$ outputs RDF triples $T = \{(s, p, o)\}$ that accurately describe the scene under O.

Given synchronized sensor readings S_t and ontology O, our goal is to compute:

$$T_t = f_{\text{reason}}\big(f_{\text{fuse}}(S_t),\, O\big). \tag{2}$$

$$\text{such that} \begin{cases} \text{Latency}(T_t) \le 200 \text{ ms}, \\ \text{Bandwidth}(T_t) \le 2 \text{ MB/frame}, \\ \text{Accuracy}(T_t) \text{ is maximized}, \\ T_t \subseteq O \quad \text{(closed-world semantic consistency)}. \end{cases}$$

This formulation highlights the fundamental trade-off between data/computation resources and semantic accuracy, motivating our sytem's dynamic, ontology-guided fusion approach.

B. AV Side

We project raw LiDAR point clouds into a BEV grid by first filtering points to a cuboidal region of interest (ROI) and then discretizing their 3D coordinates into a 2D representation that preserves spatial structure while enabling vision-style processing.

From the BEV, we detect objects using an object detector, producing $\{(x_j^{\min}, y_j^{\min}, x_j^{\max}, y_j^{\max}, c_j)\}_{j=1}^{M}$, where M is the number of detections and c_j the class label. The detections are rendered onto the BEV and paired with RGB inputs for the AI agent, improving robustness under occlusions and environmental uncertainty.

To adapt LiDAR usage, we use discrete confidence thresholds $\{\tau_1, \ldots, \tau_k\} \subset [0, 1]$ with $\tau_1 > \cdots > \tau_k$ and associated sampling rates $\gamma_1 < \cdots < \gamma_k$ ($\gamma_i \in (0, 1]$). At time t, RGB is always used, while LiDAR is sampled at rate $r_t = \gamma_i$ when semantic confidence falls below the corresponding threshold, trading sensor redundancy against compute and bandwidth.

C. The Edge-Server Side

Our system adaptively invokes LiDAR at the edge using semantic confidence, closing a loop of *perception* → *reasoning* → *perception*. Confidence is derived from ontology-aligned RDF triples predicted by an LVLM and is used to modulate sensor usage.

The Ontology Summarization Module distills a TTL ontology O into a compact context S that is prepended to each LVLM prompt to enforce domain constraints.

For each RGB batch I^b_{rgb}, the Adaptive Prompt Engine queries the LVLM:

$$(P^b_{\text{rgb}}, C^b_{\text{rgb}}) \leftarrow \texttt{QueryLVLM}(I^b_{\text{rgb}}, S, DefaultPrompt),$$

and normalizes each confidence by dividing it by the weather factor w_{weather} (e.g., Day/Night/Foggy weights 1.00/1.25/1.30).

If $\bar{c}_t^{\text{adj}} < \tau$, it requests LiDAR bird's-eye-view (BEV) I^b_{lidar} and performs fusion:

$$(P^b_{\text{fusion}}, C^b_{\text{fusion}}) \leftarrow \texttt{QueryLVLM}(I^b_{\text{rgb}} \cup I^b_{\text{lidar}}, S, CustomPrompt),$$

otherwise it uses P^b_{rgb}. In both cases, valid triples are entity-linked, ontology-aligned, and emitted for V2X sharing, enabling low latency when confidence is high and robust reasoning when it is low (no fine-tuning required).

IV. Experimental Evaluation

A. Setup

We evaluate our proposed approach on the 3CSim dataset [1] in CARLA [12], restructured in a KITTI-like format [13], using 1920×1080 RGB frames (926 KB) and LiDAR scans (31k points, 1.1 MB). Each scan is converted to a 563×563 three-channel BEV image (564 KB) and processed by a Faster R-CNN detector (Section III-B) trained on 2,000 labeled CARLA samples (rotation ignored). GPT-4 used for inferencing.

Fig. 2. Multimodal perception under dense fog: (Left) RGB view; (Right) LiDAR BEV detection showing robust object localization.

We consider three weather conditions (Clear, Foggy, Rainy), two times of day (Day, Night), and three corner-case types (Behavior, Evidence, State), yielding 18 test cases with 10 scenarios each (180 total) as shown in Fig. 2. We compare Camera-Only, LiDAR-Only, static fusion baselines, and zero-context LVLMs using: (i) semantic F_1 vs. gold triples, (ii) bandwidth (MB/frame), and (iii) for LVLM baselines, inter-model consistency (mean pairwise Jaccard similarity of triple sets). We define five fixed fusion plans (A-E) and report, per plan, the mean adjusted confidence $\bar{c}_t^{\text{adj}}$, semantic F_1 (vs. gold triples), and bandwidth (MB/batch). We compare against our framework by sweeping $\tau \in [0.60, 0.95]$.

B. Fixed Plan Performance

To quantify how different fixed RGB:LiDAR mixes trade semantic accuracy against data cost, we evaluated five plans (A–E, batch size=5) and measured for each plan both the semantic F_1 score and the bandwidth per batch. Fig. 3 shows F_1 plotted against bandwidth, with each bar annotated by its plan ratio (RGB:LiDAR) and measured MB/frame.

Plans B (4:1) and C (3:2) achieve the highest F_1 scores (0.92 and 0.90) while remaining under 6 MB per batch. By contrast, the camera-only extreme (Plan A, 5:0) reaches only F_1=0.80 at 4.60 MB, and the LiDAR-only extreme (Plan E, 0:5) falls to F_1=0.52 despite incurring 7.00 MB. This clearly shows that adding just one or two LiDAR frames yields the bulk of the accuracy gain, after which marginal returns diminish. These results directly motivate our confidence-driven, dynamic fusion strategy .

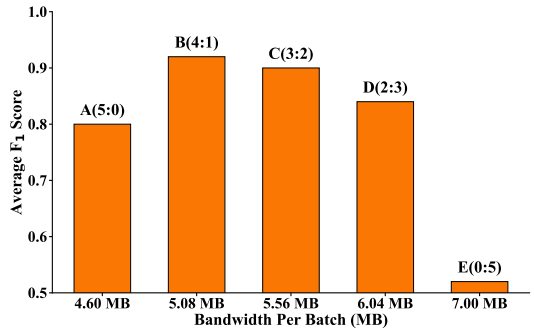


Fig. 3. Average F_1 vs bandwidth per batch for fixed RGB:LiDAR plans deliver best accuracy (<6 MB), outperforming both camera or LiDAR only.

C. Dynamic Fusion Trade-Off

Fig. 4 shows the impact across ten corner cases. Each pair of bars compares the semantic F_1 before and after applying our adaptive RGB–LiDAR fusion using plan B. Our framework consistently boosts F_1 on every scenario, most dramatically under low-visibility or high-occlusion cases (e.g., `Fallen Tree`, `Falling Barrel`, etc.), demonstrating the robustness of our confidence driven fusion strategy.

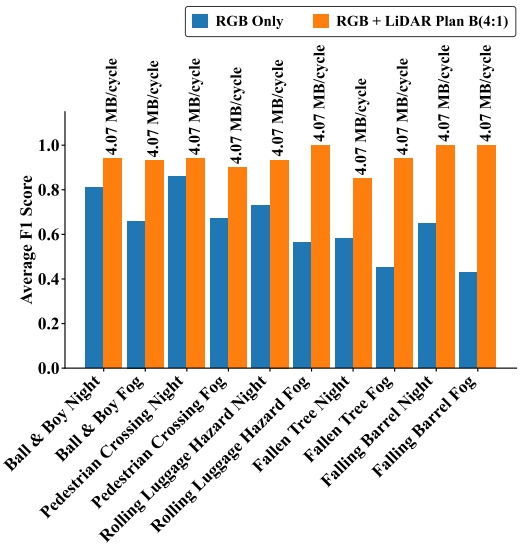


Fig. 4. Average F1 for RGB-only vs. adaptive RGB+LiDAR across 10 corner cases in fog and night.

V. Conclusion

We presented a real time corner case reasoning framework that integrates RGB and LiDAR data at the edge with a domain specific ontology and a pretrained vision language model. The framework generates ontology consistent RDF triples for semantic V2X knowledge sharing and extends retrieval augmented generation toward multimodal, constraint guided reasoning in autonomous driving.

Future research will investigate cooperative perception across multiple vehicles to mitigate the limitations of single vehicle observations, particularly in occluded regions. By leveraging communication among nearby vehicles and combining complementary viewpoints, the system can construct a richer and more comprehensive representation of the environment, enabling scene understanding at the level of a vehicle cluster rather than from an isolated vehicle perspective.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic, and the Slovak Academy of Sciences under Grant VEGA 1/0685/23 and by the Slovak Research and Development Agency under Grant APVV SK-CZ-RD-21-0028 and APVV-23-0512.

References

[1] M. Čavojsky, E. Slapak, M. Dopiriak, G. Bugar, and J. Gazda, "3CSim: CARLA Corner Case Simulation for Control Assessment in Autonomous Driving," in *2024 IEEE 8th International Conference on Information and Communication Technology (CICT)*, 2024, pp. 1–6.

[2] J.-A. Bolte, A. Bar, D. Lipinski, and T. Fingscheidt, "Towards Corner Case Detection for Autonomous Driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 438–445.

[3] M. Nawaz, J. K.-T. Tang, K. Bibi, S. Xiao, H.-P. Ho, and W. Yuan, "Robust Cognitive Capability in Autonomous Driving Using Sensor Fusion Techniques: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3228–3243, 2024.

[4] Y. Tian, K. Wang, Y. Wang, Y. Tian, Z. Wang, and F.-Y. Wang, "Adaptive and azimuth-aware fusion network of multimodal local features for 3D object detection," *Neurocomputing*, vol. 411, pp. 32–44, 2020.

[5] D. Fu, X. Li, L. Wen, M. Dou, P. Cai, B. Shi, and Y. Qiao, "Drive Like a Human: Rethinking Autonomous Driving with Large Language Models," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2024, pp. 910–919.

[6] Z. Ma, B. VanDerPloeg, C.-P. Bara, H. Yidong, E.-I. Kim, F. Gervits, M. Marge, and J. Chai, "DOROTHIE: Spoken Dialogue for Handling Unexpected Situations in Interactive Autonomous Driving Agents," 2022. [Online]. Available: https://arxiv.org/abs/2210.12511

[7] A. Elhafsi, R. Sinha, C. Agia, E. Schmerling, I. Nesnas, and M. Pavone, "Semantic Anomaly Detection with Large Language Models," 2023. [Online]. Available: https://arxiv.org/abs/2305.11307

[8] I. Tulleners, T. Moers, T. Schulik, and M. Sedlacek, "A Semi-Automated Corner Case Detection and Evaluation Pipeline," 2023. [Online]. Available: https://arxiv.org/pdf/2305.16369

[9] K. Li, K. Chen, H. Wang, L. Hong, C. Ye, J. Han, Y. Chen, W. Zhang, C. Xu, D.-Y. Yeung, X. Liang, Z. Li, and H. Xu, "CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving," 2022. [Online]. Available: https://arxiv.org/abs/2203.07724

[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, ..., and S. Riedel, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.

[11] M. Luti, D. Feldman, Z. Hu, S. Mehrotra, M. Mendoza, N. Venkatasubramanian, R. Yus, and R. Eguchi, "FOVDA: A Federated Architecture for Overcoming Data Silos in Water Domain [Vision]," in *Future of Information and Communication Conf.* Springer, 2025, pp. 625–646.

[12] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 78. PMLR, 2017, pp. 1–16.

[13] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354–3361.

Sb-Based Solders for Large-Area Joints in Power Electronics: Reflow Profile Optimization

[1]*Peter NEMERGUT (2nd year)*
Supervisor: [2]Alena PIETRIKOVÁ

[1,2]Dept. of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]peter.nemergut@tuke.sk, [2]alena.pietrikova@tuke.sk

***Abstract*—This paper addresses the optimization of the reflow soldering process for large-area joints in power electronics, focusing on an Sn95Sb5 solder preform used in power modules between a DBC substrate and a baseplate. The Sn95Sb5 alloy represents a promising lead-free alternative for high-temperature applications due to its higher melting temperature compared with standard SAC alloys and its improved intermetallic compound (IMC) stability. Experimentally, two reflow profiles were compared, differing in peak temperature and time above liquidus (TAL). Joint quality was evaluated by scanning acoustic microscopy (SAM). The results show that extending TAL together with increasing peak temperature led to an increase in the average void area (from 0.82% to 1.82%) and in the biggest void area (from 0.29% to 0.68%), highlighting the need for precise control of the reflow process. The discussion further points to solder non-uniformity (variable bondline thickness) caused by baseplate pre-warping. Future work will focus on Sb-based reinforced preforms (InFORMS®) to improve bondline thickness and uniformity.**

***Keywords*—antimony-based solder, large-area joint, optimization, power electronics, reflow process**

I. Introduction

In power electronics, increasingly stringent requirements are being placed on the reliability and thermal management of power modules. Large-area solder joints between the DBC (Direct Bonded Copper) substrate and the baseplate (BP) represent critical locations, particularly with respect to thermal resistance and thermo-mechanical loading under power and thermal cycling. Conventional lead-based solders have historically been preferred in these applications due to their robust behavior and wide process tolerance. However, the transition toward lead-free solutions necessitates the identification of alloys offering a higher temperature margin and stable microstructural characteristics. This work focuses on an Sb-based (antimony) solder alloy (Sn95Sb5) in the form of a preform, which is suitable for large-area joints and enables a well-defined solder thickness within the joint.

The objective of this paper is to optimize the reflow profile for a large-area joint between a DBC substrate and a BP formed using a Sn95Sb5 solder preform, and to experimentally assess the influence of peak temperature and TAL on voiding and bondline thickness non-uniformity in the large-area joint. Joint quality is evaluated non-destructively using scanning acoustic microscopy (SAM), which enables quantification of both the total voided area and the presence of dominant local voids [1], [2], [3].

II. Sb-Based Solders as a Substitute for Lead Solders

Sb-based solder alloys are attractive for power-electronic applications primarily due to their higher melting temperature and the formation of a stable Sn–Sb intermetallic phase within the Sn matrix. The Sn95Sb5 alloy (237–240 °C) provides a larger temperature margin than conventional SAC solders (217–221 °C), thereby improving joint robustness at elevated operating temperatures. Sb partially dissolves in Sn, contributing to solid-solution strengthening, while the remaining fraction forms finely dispersed Sn–Sb particles. This dispersion promotes microstructural refinement and contributes to the overall stability of the soldered joint. From a practical perspective, the thermal conductivity of Sn95Sb5 (≈64.7 W/m·K) is also relevant for large-area joints, as the solder layer directly affects the thermal resistance of the overall power-module stack [2], [3], [4].

Using a solder preform allows the solder volume to be precisely defined, which improves process repeatability for large-area joints. Compared with solder paste deposition, a preform mitigates issues related to non-uniform solder thickness and facilitates achieving a target thickness and higher homogeneity of the solder layer (bondline). However, the use of a preform alone does not guarantee bondline uniformity, particularly in the presence of BP warpage. Local non-coplanarity of the joining surfaces or insufficient clamping force during reflow can lead to molten solder redistribution, typically resulting in solder depletion at the joint periphery and accumulation in the central region. Such non-uniformity is critical because it can increase local mechanical stresses and alter the conditions for void nucleation and coalescence or even promote crack initiation within the solder joint.

This issue is particularly relevant for high-temperature Sb-based solders, which are more sensitive to bondline thickness variations and local stress concentrations due to their increased stiffness and limited stress-relaxation capability. Eliminating bondline non-uniformity in large-area joints is therefore an important practical challenge, typically addressed through reflow process optimization (peak temperature, TAL, vacuum or reducing stages, and applied pressure). The motivation is to analyze how reflow parameters influence bondline formation when using Sb-based preforms and to identify process windows that lead to stable, long-term reliable large-area joints for power electronics [4].

III. Reflow Process Optimization

The aim of the optimization is to define a reflow profile for a large-area joint between the DBC substrate and BP using an Sn95Sb5 preform, such that voiding is minimized while limiting the adverse effects of prolonged time in the liquid phase (molten solder redistribution, IMC growth, and the risk of local defects). Two profiles were compared, differing in peak temperature and time above liquidus.

A. Description of the Samples and Assembly

The samples consisted of a Cu BP (104 x 59 x 3 mm) with a Ni surface finish. Four Sn95Sb5 solder preforms with a thickness of 200 µm were placed on the BP, and four DBC substrates (38 x 26 x 0.96 mm) were positioned on top of the preforms (Fig. 1 and Fig. 3). Each DBC substrate was loaded with 20 g dead weight. Joint quality was evaluated non-destructively using SAM, specifically *void area per DBC area* and *the biggest void area*.

B. Reflow Profile Parameters

The samples were processed by reflow in a static laboratory furnace using a formic-acid reducing atmosphere (HCOOH). At the beginning of the process, two consecutive vacuum cycles (3 mbar) were applied in combination with N_2. Subsequently, vacuum was re-activated after the pre-soak phase during ramp-up to the peak temperature and maintained continuously throughout the entire peak stage. Heating was performed at 30 °C/s (measured on the sample) up to a pre-soak temperature of 190 °C. At 190 °C, HCOOH was applied via a bubbler (10 slm for 210 s), followed by evacuation and subsequent heating to the peak temperature according to the specific sample profile. After the peak stage, the chamber was backfilled with N_2 (20 slm). Cooling was controlled with an average rate of 0.4 °C/s. Within the optimization study, peak temperature and TAL (defined as the time above 240 °C) were varied simultaneously due to the setting limitations of the laboratory furnace: sample A - peak 255 °C (≈ +15 °C above liquidus), TAL 60 s; sample B - peak 280 °C (≈ +40 °C above liquidus), TAL 180 s [1], [5].

C. Result and Interpretation

Table I indicates an increase in voiding at higher peak temperature and longer TAL. For sample A (Fig. 2), the average void area per DBC area was 0.82% (max. 1.32%), and the biggest void area was 0.29%. For sample B (Fig. 4), the average void area per DBC area increased to 1.82% (max. 2.51%), and the biggest void area increased to 0.68%. Local exceedance of the 2% level was also observed, which may increase the effective thermal resistance and local overheating of the power module.

TABLE I
Evaluation of the Void Area per DBC and Biggest Void Area

Sample	DBC	Void area per DBC area (%)	Average void area (%)	Biggest void area (%)	Max. biggest void area (%)
A	1	0.88	0.82	0.16	0.29
	2	1.32		0.29	
	3	0.41		0.12	
	4	0.65		0.12	
B	1	1.30	1.82	0.18	0.68
	2	2.51		0.27	
	3	1.86		0.29	
	4	1.60		0.68	

The results suggest that excessively long TAL can promote void coalescence and growth and can increase the probability of dominant local cavities, as reflected by the increase in the "biggest void" metric. Practically, this implies that even under vacuum and/or reducing reflow conditions, extending the time in the liquid phase should not be applied by default. A shorter TAL with controlled transitions is preferable to support outgassing and wetting while avoiding unnecessary exposure of the molten solder to elevated temperature.

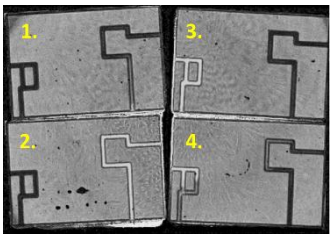


Fig. 1. Placement of DBC substrates on the BP (sample A)

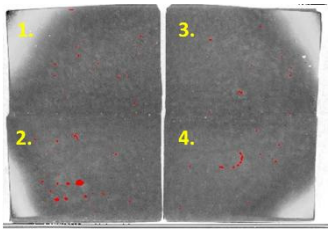


Fig. 2. SAM analysis - void content evaluation for sample A (red regions)

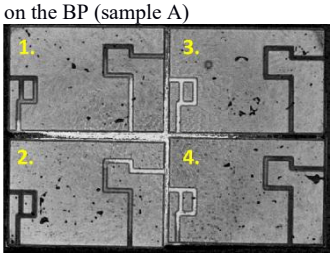


Fig. 3. Placement of DBC substrates on the BP (sample B)

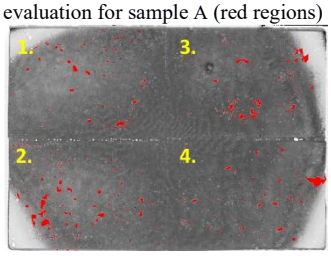


Fig. 4. SAM analysis - void content evaluation for sample B (red regions)

The SAM images (Fig. 2 and Fig. 4) further reveal non-uniform solder distribution, typically characterized by reduced thickness at the joint edges (lighter contrast) and accumulation in the central region (darker contrast). This behavior is consistent with the influence of baseplate warpage and locally varying standoff during reflow. A longer TAL may further intensify this molten solder redistribution. Therefore, reflow optimization should consider not only the void fraction but also bondline uniformity, as it directly affects local thermal resistance and internal stresses in the solder joint.

IV. Recommendation for Practice

Based on comparison of the results of the SAM analysis of samples, it is evident that for a large-area joint between the DBC substrate and BP using an Sn95Sb5 preform, increasing the peak temperature or extending TAL is not an effective approach to minimize voiding. Instead, the reflow process should be optimized to promote reliable wetting and efficient gas release while limiting void coalescence and excessive molten solder redistribution within the joint. From a manufacturing perspective, the presented results provide a practical basis for defining a robust reflow process window for Sb-based preforms, with reduced sensitivity to typical production variability and improved process stability for repeatable fabrication of large-area power modules.

With respect to voiding, it is advantageous to keep TAL at the minimum value that still ensures complete reflow across the entire joint area and uniform wetting of the interfaces. In this study, the shorter TAL (60 s) combined with a lower peak temperature (255 °C) resulted in both a lower average void fraction and a reduced occurrence of dominant voids. In contrast, the longer TAL (180 s) at a higher peak temperature

(280 °C) increased the overall voiding level and raised the probability of large local cavities, which are more critical for heat transfer than uniformly dispersed small voids.

For large-area assemblies, it is also essential to limit temperature gradients between the center and edges of the BP. Non-uniform overheating can lead to different effective times in the liquid phase across the joint area, promoting uneven molten solder flow, local gas entrapment, and reduced bondline uniformity. Practically, this requires careful setting of the pre-soak stage and heating rates so that the temperature of the entire assembly stabilizes before reaching liquidus.

Mechanical boundary conditions during reflow significantly affect large-area joint quality, particularly the applied pressure and the planarity of the assembly. BP warpage leads to local variations in standoff and promotes molten solder redistribution toward the joint center, causing reduced solder thickness at the edges. This effect is amplified at longer TAL. Consequently, reflow profile optimization is only meaningful when combined with control of mechanical fixturing, since conventional low-temperature SAC solders may be limiting for such applications in terms of mechanical and thermo-mechanical stability.

V. Antimony Solder For Achieving Uniform Solder Bondline Control

A promising approach to mitigate solder non-uniformity is the use of reinforced preforms such as InFORMS®, which incorporate an integrated reinforcement matrix and are designed to provide more consistent bondline thickness while reducing stress concentrations caused by assembly tilt and/or lack of planarity. According to the InFORMS® product datasheet, the material is a reinforced, matrixed solder composite in which the reinforcement remains within the joint after reflow process. This reinforcement is intended to increase mechanical robustness (improved resistance to creep and fatigue) while preserving the intrinsic properties of the selected solder alloy. In addition, it improves process repeatability and bondline uniformity, which directly contributes to enhanced thermal and mechanical joint reliability. The manufacturer also reports the use of InFORMS® specifically for “bonding the DBC to the baseplate” in IGBT module applications [6], [7].

In the next phase of this research, we plan to evaluate InFORMS® as an alternative to standard, non-reinforced preforms without requiring major modifications to the reflow process. The assessment will focus on bondline uniformity, including center-to-edge differences and the statistical distribution of bondline thickness, based on metallographic cross-sections and 3D profilometry. In parallel, the correlation between bondline uniformity and void characteristics will be analyzed, as well as the impact of bondline thickness on delamination and failure initiation under thermo-mechanical loading. Attention will also be paid to process stability under a defined clamping force and a known level of BP pre-warping (warpage), evaluated before and after the reflow process.

VI. Conclusion

This paper assessed the suitability of high-temperature Sn95Sb5 solder in preform form for a large-area joint between a DBC substrate and a BP and analyzed the influence of reflow profile settings on joint quality with respect to voiding and solder layer homogeneity. A comparison of two reflow profiles showed that increasing the peak temperature and significantly extending TAL results in higher overall voiding as well as a higher occurrence of dominant local cavities compared with the shorter TAL profile.

The results confirm the need to define a process window that minimizes TAL while maintaining reliable wetting and outgassing, and that is robust against manufacturing variability (thermal loading of the assembly, non-coplanarity, and clamping-force fluctuations). Optimization should therefore address not only temperature selection, but also the control of temperature gradients and stress states in the large-area joint.

For long-term module reliability, it is necessary to evaluate not only the total void percentage but also the biggest void, the spatial distribution of voids, and particularly bondline uniformity. The observed solder non-uniformity caused by BP warpage and variable standoff can lead to local differences in thermal resistance and to mechanical stress concentrations. Future work will focus on validating Sb-based reinforced preforms (InFORMS®) to stabilize the bondline, broaden the robust process window, and improve manufacturing repeatability without major changes to the existing reflow process.

Acknowledgment

The Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00002.

References

[1] M. Chnapko, J. Sitar, M. Frivaldsky, and L. Hargas, “The Power Electronic Soldering Process: An Evaluation of Soldering Materials and Basic Soldering Principles,” *Applied Sciences*, vol. 15, no. 14, Jul. 2025, doi: 10.3390/app15147732.

[2] A. A. El-Daly, A. Fawzy, A. Z. Mohamad, and A. M. El-Taher, “Microstructural evolution and tensile properties of Sn–5Sb solder alloy containing small amount of Ag and Cu,” *Journal of Alloys and Compounds*, vol. 509, no. 13, pp. 4574–4582, Mar. 2011, doi: 10.1016/j.jallcom.2011.01.109.

[3] T. Kobayashi, I. Shohji, and Y. Nakata, “Effect of Power Cycling and Heat Aging on Reliability and IMC Growth of Sn-5Sb and Sn-10Sb Solder Joints,” *Advances in Materials Science and Engineering*, vol. 2018, no. 1, p. 4829508, 2018, doi: 10.1155/2018/4829508.

[4] G. Zeng, S. McDonald, and K. Nogita, “Development of high-temperature solders: Review,” *Microelectronics Reliability*, vol. 52, no. 7, pp. 1306–1322, Jul. 2012, doi: 10.1016/j.microrel.2012.02.018.

[5] L. Ji *et al.*, “Tin fog effect analysis during fluxless soldering under reflow with formic acid: Methodology, mechanism, and reliability impact,” *Journal of Materials Research and Technology*, vol. 36, pp. 5366–5380, May 2025, doi: 10.1016/j.jmrt.2025.04.151.

[6] “Reinforced Preforms,” Indium Corp. Accessed: Feb. 13, 2026. [Online]. Available: https://www.indium.com/products/solder-preforms/reinforced-preforms/

[7] “Achieving Uniform Solder Bondline Control Between Substrate and Baseplate in IGBTs,” Indium Corp. Accessed: Feb. 13, 2026. [Online]. Available: https://www.indium.com/blog/achieving-uniform-solder-bondline-control-between-substrate-and-baseplate-in-igbts-part-i/

Contextual Gamified Cybersecurity Education

[1]*Lenka BUBEŇKOVÁ (2nd year),*
Supervisor: [2]*Emília PIETRIKOVÁ*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]lenka.bubenkova@tuke.sk, [2]emilia.pietrikova@tuke.sk

***Abstract*—Cybersecurity incidents persistently exploit the human factor despite expanding training and regulatory obligations. This study evaluates three gamified training modalities: rapid phishing recognition, scenario-based investigation, and contextual incident-driven gameplay. Findings show modality-specific behavioral effects, with improvements in reflex detection, situational awareness, and holistic risk perception. The results support adaptive, behavior-focused cybersecurity education aligned with regulatory requirements.**

***Keywords*—Cybersecurity education, gamification, human factor, behavioral training, regulatory frameworks**

I. Introduction

Cybersecurity has evolved from a peripheral technical concern into a global societal priority. Ransomware campaigns against critical infrastructure, persistent phishing waves, and supply-chain breaches reveal that vulnerabilities are not purely technical: human decision-making remains a primary attack surface despite advances in endpoint protection, detection, and authentication mechanisms. This pattern is consistently reflected in large-scale monitoring. European Union Agency for Cybersecurity (ENISA) identifies phishing and social engineering among the most prevalent cross-sector threats and continues to emphasize the role of human error in successful incidents [1]. Similarly, the Verizon Data Breach Investigations Report (DBIR) reports that nearly 60% of incidents involve a human element, including phishing interaction, credential compromise, and procedural bypassing [2].

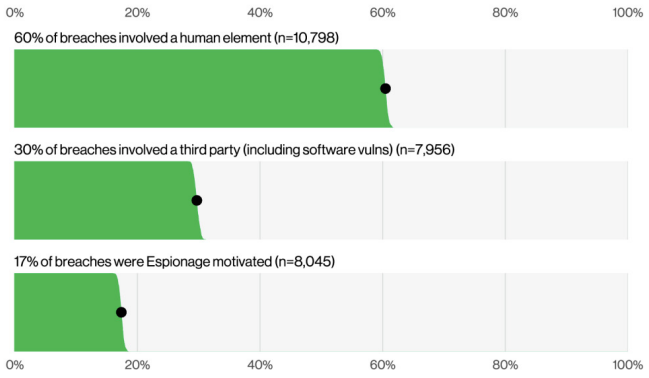


Fig. 1. Selected characteristics of security incidents: proportion of incidents involving the human element, third parties, and espionage motivation.

The persistence of human-driven breaches highlights a key paradox: awareness efforts and training completion rates increase, yet secure behavior does not improve proportionally. Evidence suggests that knowledge transfer alone is insufficient, as users act under cognitive biases, urgency cues, trust assumptions, and social pressure, applying security knowledge inconsistently [3]. Empirical research further indicates that susceptibility to phishing is associated with risk perception and individual differences, rather than with cybersecurity literacy [4], [5], [6]. This motivates the need for training interventions that evaluate behavioral change under realistic conditions and capture outcomes beyond short-term knowledge tests.

Cybersecurity education is increasingly formalized by EU regulatory frameworks. Network and Information Security Directive 2 (NIS2) and Digital Operational Resilience Act (DORA) embed awareness, training, and exercises into organizational risk governance and operational resilience [7], [8]. Cyber Resilience Act (CRA) reinforces cybersecurity-by-design and lifecycle vulnerability handling for products with digital elements, indirectly increasing the demand for secure development skills [9]. The AI Act adds requirements for human oversight and robustness in high-risk AI systems, raising the importance of mitigating automation bias and improving decision-making under uncertainty [10]. However, while these frameworks mandate training, they provide limited guidance on measuring effectiveness beyond formal completion, leaving a gap in standardized behavioral metrics and longitudinal evaluation. To address the identified research gaps, this study is guided by the following research questions:

RQ1: How do different gamified training modalities influence behavioral responses to cybersecurity threats?

RQ2: Do rapid-response and scenario-based training approaches lead to measurable improvements in decision-making patterns?

RQ3: Can contextual, narrative-driven gameplay enhance holistic risk perception beyond isolated threat recognition?

II. Introduction to the Research Problem

The analysis of reviewed publications reveals three structural limitations that constrain the systematic application of gamification in cybersecurity education. First, most studies adopt a short-term evaluation perspective. Training effectiveness is typically measured immediately after intervention, without longitudinal evidence assessing retention of secure behavioral patterns [11], [12]. As a result, the persistence of behavioral change remains largely unverified. Second, research samples are frequently limited in size and diversity. Many studies rely on homogeneous populations, most commonly IT students [13], [14]. This restricts generalizability to broader user groups, including non-technical employees, children, and older

TABLE I
OVERVIEW OF RESEARCH GAPS AND STUDY LIMITATIONS IN REVIEWED PUBLICATIONS

Source	Research Gaps	Study Limitations
All et al. [13]	No unified evaluation methodology for DGBL effectiveness.	Lack of pre-tests; extremely small samples.
Arachchilage & Love [15]	Limited focus on home end-users in phishing education.	Narrow age group; no URL redirection scenarios.
Ros et al. [18]	Unclear impact of specific game elements on success perception.	Quasi-experimental design; metaphor ambiguity.
Fink et al. [17]	Limited research on situational awareness in cyber defense.	Competitive bias; subjective evaluation.
Faith et al. [19]	Applicability of gamification across expertise levels unclear.	Ongoing research.
Putz et al. [11]	Lack of longitudinal retention studies.	Knowledge-focused; culturally limited sample.
Zainuddin et al. [12]	Weak theoretical grounding of gamification–motivation link.	Short intervention duration.
Turchi et al. [14]	Underexplored role of gameplay in computational thinking.	Very small sample.
Witeck et al. [16]	Limited validation of higher Bloom levels via gamification.	Subjective evaluation.

adults [15]. Third, the literature demonstrates methodological fragmentation. Evaluation approaches vary widely—from knowledge acquisition to motivation or behavioral performance—without standardized metrics or comparable baselines [13], [16]. The absence of harmonized evaluation frameworks complicates cross-study comparison and risks overstating effectiveness. Additional gaps remain in situational awareness, gameplay design, and applied cyber defense contexts [17], [18], [19]. Collectively, these limitations motivate the need for large-scale, behavior-focused training experiments that combine multiple gamified modalities and evaluate outcomes beyond short-term knowledge acquisition.

III. ONGOING EXPERIMENTAL TRAINING DEPLOYMENTS AND BEHAVIORAL OBSERVATIONS

To address previously identified limitations—short-term evaluation, homogeneous samples, and fragmented behavioral metrics—an ongoing series of large-scale cybersecurity education workshops was conducted. The experimental deployments were designed to extend beyond knowledge-acquisition metrics and to capture behavioral adaptation, situational reasoning, and decision-making patterns across varying instructional modalities. The training design integrates three gamified interventions targeting distinct cognitive dimensions: rapid phishing recognition, scenario-based investigation, and contextual cyber risk gameplay. Collectively, these interventions were deployed across several hundred participants spanning multiple educational environments, enabling comparative cross-modal analysis and contributing to one of the largest behavioral datasets. The experimental design was structured to address the defined research questions by comparing behavioral responses across three distinct training modalities.

A. Rapid Recognition Game: Phishing Detection Under Time Pressure

The first modality addressed reflex threat recognition through time-constrained phishing classification tasks. Participants completed a diagnostic pretest, followed by an interactive rapid-response game, and concluded with a structurally equivalent posttest. Average pretest performance reached 9.91 out of 12 points, while posttest scores averaged 9.86. Among participants who completed both phases, approximately one-third improved, one-third remained stable, and one-third declined. Although aggregate gains were limited, item-level analysis revealed targeted improvements in visually explicit phishing indicators and urgency-framed deception cues. Conversely, performance decreased in cognitively ambiguous scenarios involving layered or hybrid fraud signals. These findings suggest that speed-based training enhances reflex detection capabilities but may introduce cognitive overload in analytically complex threat evaluations. This aligns with the previously identified need for complementary reflective learning modalities.

B. Scenario-Based Collaborative Investigation Game

To address gaps in situational awareness development, the second intervention employed narrative-driven group investigations. Participants reconstructed cyber incident timelines by ordering events according to perceived escalation severity. Sequencing accuracy averaged approximately 49.6% within directly evaluated datasets, with broader aggregation indicating even lower baseline alignment. Risk perception scaling revealed that early compromise indicators were consistently rated as high risk (mean 3.84/5), whereas later-stage artifacts were frequently underestimated. These results reveal a structural disconnect between static phishing knowledge and understanding of dynamic attack progression. While participants recognized isolated red flags, they struggled to contextualize threats temporally within incident chains. Qualitative feedback further identified event sequencing as the most cognitively demanding activity, reinforcing situational reasoning as a critical yet underdeveloped competency in cybersecurity education.

C. Contextual Cyber Risk Investigation Game

The third modality employed an incident-driven modular game centered on contextual cybersecurity risk awareness. Participants analyzed a narrative scenario involving a fictional user and reconstructed behavioral and environmental factors contributing to a cyber incident. Unlike phishing classification tasks, this approach addressed broader risk vectors, including unsecured Wi-Fi usage, credential exposure, and post-incident response behavior. The investigative gameplay structure was organized into thematic modules reflecting real-world cyber risk domains. These included phishing and social engineering, secure web browsing and deceptive website recognition, risks associated with public Wi-Fi networks, password security and multi-factor authentication, web tracking and cookie-based monitoring, and appropriate response actions following a security incident. Rather than focusing on isolated threat identification, participants explored how individual decisions cumulatively contributed to incident escalation. The modular architecture enabled contextual adaptation across different user populations. This persona-sensitive structure directly responds to the generalization gap identified in traditional awareness

training, where content is often delivered uniformly regardless of user context. By embedding cybersecurity risks within relatable behavioral narratives, the game supported holistic threat perception, situational reasoning, and consequence mapping. The findings indicate that contextualized experiential learning may provide a critical bridge between abstract awareness and applied behavioral resilience.

D. Participant Experience and Cross-Modal Training Effects

Survey feedback complemented behavioral observations by providing insight into perceived engagement and learning value. Participants rated the rapid recognition module as highly engaging, whereas scenario-based gameplay was perceived as immersive but cognitively demanding. A dominant proportion of respondents reported improved fraud recognition capability following the interventions, and the majority expressed willingness to participate in similar training formats in the future. Reported difficulty factors aligned with performance variability, with event sequencing and time-constrained decision-making identified as the most challenging elements. Synthesizing performance metrics, perception data, and observational insights reveals distinct cross-modal behavioral effects. Speed-based detection strengthens reflex recognition but requires analytical reinforcement, while scenario investigations enhance situational awareness at the cost of higher cognitive demand. Contextual modular gameplay supports progressive internalization of broader cyber risk patterns, particularly when adapted to user-specific environments. Engagement levels across modalities confirm the strong motivational viability of gamified cybersecurity education. However, behavioral adaptation remains modality-dependent rather than universally transferable. Collectively, these findings directly address previously identified research gaps by contributing large-sample behavioral evidence, multimodal training comparisons, and expanded evaluation metrics extending beyond short-term knowledge acquisition. Ongoing deployments continue longitudinally, enabling future analysis of retention dynamics and adaptive training personalization.

IV. Conclusion

The findings further reinforce the continued relevance of contextually adaptive cybersecurity training. While phishing awareness and conceptual threat recognition demonstrate relatively high baseline success rates, the critical challenge lies not in knowledge acquisition but in the subconscious shaping of behavioral responses under real-world exposure conditions. This shifts the central research question from whether individuals recognize cyber threats to how training interventions can most effectively influence instinctive decision-making in risk situations. Contextual personalization—accounting for user roles, environments, and behavioral patterns—emerges as a necessary component of effective cybersecurity education. A parallel implication concerns the integration of behavioral insights into secure system design. Observing how users interpret, misinterpret, or respond to risk scenarios provides valuable input for developing more resilient programming practices, interface safeguards, and web security mechanisms. Security architectures are frequently described as being only as strong as their weakest human link. However, this does not preclude the proactive design of systems that better protect, guide, and compensate for user vulnerabilities. Future research will therefore focus on longitudinal behavioral tracking and on translating observed training outcomes into adaptive educational models and secure development frameworks.

Acknowledgment

This work was supported by project KEGA 061TUKE-4/2025 "Building bridges between university and high school ICT education".

References

[1] European Union Agency for Cybersecurity., *ENISA threat landscape 2024: July 2023 to June 2024*. Publications Office, 2024.

[2] Verizon, "2025 data breach investigations report," Verizon Business, Technical Report, 2025, analyzes over 22,000 real-world security incidents, including more than 12,000 confirmed data breaches across industries.

[3] L. Gallo, D. Gentile, S. Ruggiero, A. Botta, and G. Ventre, "The human factor in phishing: Collecting and analyzing user behavior when reading emails," *Computers & Security*, vol. 139, p. 103671, 2024.

[4] E. Boritz, C. Ge, and K. E. Patterson, "Factors affecting employees' susceptibility to cyber attacks," *SSRN Electronic Journal*, 2022. [Online]. Available: https://www.ssrn.com/abstract=4088873

[5] N. Beu, A. Jayatilaka, M. Zahedi, M. A. Babar, L. Hartley, W. Lewinsmith, and I. Baetu, "Falling for phishing attempts: An investigation of individual differences that are associated with behavior in a naturalistic phishing simulation," *Computers & Security*, vol. 131, p. 103313, 2023.

[6] M. Gratian, S. Bandi, M. Cukier, J. Dykstra, and A. Ginther, "Correlating human traits and cyber security behavior intentions," *Computers & Security*, vol. 73, pp. 345–358, Mar. 2018.

[7] European Parliament and Council of the European Union, "Directive (eu) 2022/2555 of the european parliament and of the council of 14 12 2022 on measures for a high common level of cybersecurity across the union," Dec. 2022, nIS2 Directive.

[8] ——, "Regulation (eu) 2022/2554 of the european parliament and of the council of 14 12 2022 on digital operational resilience for the financial sector," Dec. 2022, digital Operational Resilience Act (DORA).

[9] ——, "Regulation (eu) 2024/2847 of the european parliament and of the council on horizontal cybersecurity requirements for products with digital elements (cyber resilience act)," 2024, cyber Resilience Act.

[10] ——, "Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act)," Jun. 2024, artificial Intelligence Act.

[11] L.-M. Putz, F. Hofbauer, and H. Treiblmaier, "Can gamification help to improve education? findings from a longitudinal study," *Computers in Human Behavior*, vol. 110, p. 106392, 2020.

[12] Z. Zainuddin, S. K. W. Chu, M. Shujahat, and C. J. Perera, "The impact of gamification on learning and instruction: A systematic review of empirical evidence," *Educational Research Review*, vol. 30, p. 100326, 6 2020.

[13] A. All, E. Nunez Castellar, N. Castellar, and J. Looy, "Measuring effectiveness in digital game-based learning: A methodological review," *International Journal of Serious Games*, vol. 1, 5 2014.

[14] T. Turchi, D. Fogli, and A. Malizia, "Fostering computational thinking through collaborative game-based learning," *Multimedia Tools and Applications*, vol. 78, no. 10, p. 13649–13673, 2019.

[15] N. A. G. Arachchilage and S. Love, "A game design framework for avoiding phishing attacks," *Computers in Human Behavior*, vol. 29, no. 3, pp. 706–714, 2013.

[16] G. Witeck, A. Alves, and M. Bernardo, "Bloom taxonomy, serious games and lean learning: What do these topics have in common?" in *Learning in the Digital Era*, ser. IFIP Advances in Information and Communication Technology. Springer, Cham, 2021, vol. 610.

[17] G. Fink, D. Best, D. Manz, V. Popovsky, and B. Endicott-Popovsky, "Gamification for measuring cyber security situational awareness," in *International Conference on Augmented Cognition*, ser. Lecture Notes in Computer Science, vol. 8027. Springer, 2013, p. 656–665.

[18] S. Ros, S. González, A. Robles, L. L. Tobarra, A. Caminero, and J. Cano, "Analyzing students' self-perception of success and learning effectiveness using gamification in an online cybersecurity course," *IEEE Access*, 2023.

[19] B. F. Faith, Z. A. Long, and S. Hamid, "Promoting cybersecurity knowledge via gamification: An innovative intervention design," in *2024 Third International Conference on Distributed Computing and High Performance Computing (DCHPC)*, 2024.

A Review of Ultra-Wideband Antennas on Paper-Based Substrates

[1]*Filip MAHDAL (1st year)*
Supervisor: [2]Peter LUKACS

[1,2]Department of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]filip.mahdal@tuke.sk, [2]peter.lukacs@tuke.sk

***Abstract*— This paper presents a review of ultra-wideband (UWB) antennas implemented on paper-based substrates, with primary emphasis on dielectric material properties and their influence on antenna performance. The analysis focuses on three principal substrate categories conventional cellulose paper, aramid-based paper, and nanocellulose materials. Their relative permittivity, dielectric loss tangent, hygroscopic behavior, dielectric homogeneity, and environmental stability are compared in the context of UWB operation. Particular attention is devoted to the impact of dielectric dispersion and moisture-induced parameter variations on impedance bandwidth, resonant frequency stability, and radiation efficiency. The advantages and limitations of each substrate type are critically evaluated with respect to dielectric losses, permittivity stability, and long-term reliability. The review highlights current research directions aimed at improving dielectric consistency and environmental robustness of paper-based materials for sustainable high-frequency wireless systems.**

***Keywords*— Dielectric properties, green electronics, paper-based substrates, UWB antennas**

I. Introduction

The rapid development of wireless communication systems, wearable electronics, and Internet of Things (IoT) technologies has significantly increased the demand for low-cost, flexible, and environmentally sustainable electronic components. Among these components, antennas play a crucial role in enabling reliable wireless connectivity. Ultra-wideband (UWB) antennas, characterized by their wide operating bandwidth and low power spectral density, have attracted considerable attention due to their suitability for short-range communication, high data-rate transmission, localization systems, and sensing applications [1].

Increasing socio-political pressure to reduce emissions and environmental impact has accelerated the development of biodegradable and sustainable electronics. Conventional flexible substrates used in antenna fabrication are typically polymer-based materials, which may require hundreds of years to degrade and contribute to growing electronic waste. As a result, cellulose-based materials have emerged as promising alternatives, particularly paper substrates, which offer biodegradability, recyclability, and widespread availability [2].

Traditional antenna fabrication relies on rigid substrates such as FR4 or flexible polymer materials including polyimide and polyethylene terephthalate (PET). While these materials provide stable electrical performance, their environmental footprint has motivated research into eco-friendly substitutes. Paper-based substrates have therefore gained increasing attention for flexible RF applications due to their low cost, mechanical flexibility, compatibility with printing technologies, and reduced environmental impact. However, challenges such as higher dielectric losses, moisture sensitivity, and variability of material properties must be addressed [3], [4]. This paper presents a review of ultra-wideband antennas fabricated on paper substrates, focusing on current research trends, material properties, performance limitations, and future directions toward sustainable printed RF systems.

II. Paper-based substrates

Paper substrates represent a promising alternative to conventional dielectric materials in the field of UWB antennas, particularly in the context of low-cost, flexible, and environmentally sustainable electronics. However, their electromagnetic, mechanical, and technological properties strongly depend on the specific type of paper, its composition, and the manufacturing process. The following section systematically categorizes and discusses the principal types of paper substrates used in UWB applications [3], [5].

A. Cellulose-Based Papers

Paper generally contains a biopolymer known as cellulose within its structure. However, the mechanical and electrical properties of different types of paper vary due to differences in material composition and manufacturing processes. It is difficult to generalize a single set of properties for paper, as the content and type of its constituent components differ among various paper grades. Each type of paper serves a specific purpose or function determined by its structure and composition [3].

From an electromagnetic perspective, the relative permittivity of conventional cellulose-based papers typically ranges from $\varepsilon_r \approx 1.6 – 3.5$ at microwave frequencies, while the dielectric loss tangent $\tan\delta$ may reach values between 0.01 and 0.08, depending on fiber density and moisture content. The hygroscopic nature of cellulose represents one of the primary limiting factors for UWB applications, as variations in humidity lead to noticeable shifts in resonant frequencies, degradation of impedance matching, and changes in both relative permittivity and dielectric loss [6], [7].

From a technological perspective, one advantage of cellulose-based papers is their porous surface, which promotes good mechanical adhesion of conductive pastes. However, this porosity also allows the printed material to penetrate into the depth of the substrate, potentially reducing the effective

conductivity of antenna structures and increasing surface roughness. Porosity can be mitigated by applying protective coatings to the paper surface. In addition, functional materials may be introduced into the pores to tailor the dielectric and mechanical properties of the substrate [8].

B. Aramid-Based Paper Substrates

One approach to overcoming the unfavorable properties of cellulose-based paper involves partial or complete substitution with aromatic polyamide fibers, commonly known as aramid or Kevlar. Aramid materials are characterized by high thermal resistance, flame retardancy, and excellent tensile strength. Among the most widely recognized aramid-based paper products are DuPont Nomex materials, which represent a technologically advanced class of paper substrates intended for demanding applications [9].

Compared to cellulose-based papers, aramid papers exhibit lower hygroscopicity, higher mechanical stability, improved environmental durability, and significantly better thermal resistance. The relative permittivity of aramid papers typically ranges around $\varepsilon_r \approx 2.6 – 3.1$ at microwave frequencies. While the dielectric loss tangent may reach values between 0.02 and 0.04 [9], [10].

These properties enable the fabrication of more stable antenna structures with reduced sensitivity to environmental conditions. The disadvantages of aramid papers include higher cost, limited availability, and potentially weaker adhesion of conductive pastes without additional surface treatment [10].

The structural components of aramid paper consist of fibers that can be divided into two subcategories: flocs and fibrids. Aramid flocs are short segments of long aramid fibers, approximately 6 mm in length, with diameters around 10 μm. Aramid fibrids have an irregular shape, lengths ranging from 0.2 to 1 mm, and thicknesses below 1 μm. Fibrids act as a binding agent that holds the aramid fibers together and also serve as a filler within the aramid paper structure. The adhesion force between aramid flocs and aramid fibrids is approximately five times weaker than the force acting between fibrids themselves. The adhesion between fibrids is stronger because their contact surface area is larger than that of the fibers [9]. Fig. 1 illustrates types of fibers in aramid paper.

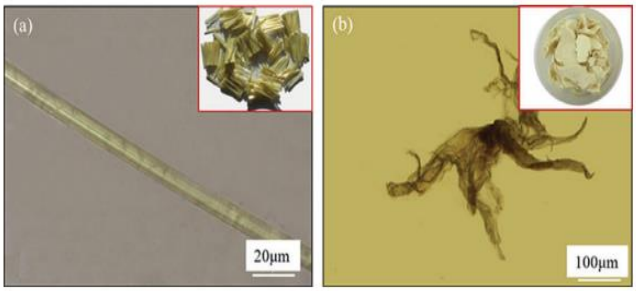


Fig. 1. Morphologies of aramid fiber: (a) flocs; (b) fibrids. [9]

C. Nanocellulose Papers

Nanocellulose papers represent an advanced class of cellulose-based substrates whose microstructure is composed of fibers or fibrils with nanometer scale dimensions. This fundamentally distinguishes them from conventional cellulose papers, which consist of micrometer-scale fibers and exhibit a highly porous structure. Owing to their more compact and homogeneous internal structure, nanocellulose materials demonstrate lower surface roughness, higher density, and improved permittivity uniformity, which is particularly important for high-frequency applications. In addition, these materials may be transparent or semi-transparent, expanding their potential use in optoelectronic, flexible RF systems, and wearable electronics [11], [12]. The relative permittivity typically ranges approximately from $\varepsilon_r \approx 3 – 4.4$ at microwave frequencies, while maintaining a low dielectric loss tangent at values between 0.01 and 0.05. The reduced porosity and lower volume fraction of air voids contribute to decreased dielectric losses compared to conventional cellulose-based papers [13], [14].

These properties translate into higher radiation efficiency and more stable impedance characteristics.

Compared to aramid papers, nanocellulose does not achieve the same level of thermal resistance; however, it offers lower dielectric losses and improved environmental compatibility due to its biodegradable nature.

Nanocellulose papers are classified according to their morphology and fabrication method into substrates based on cellulose nanofibrils (CNF), which consist of long, flexible fibrils with a high aspect ratio and provide good mechanical flexibility, and cellulose nanocrystals (CNC), which exhibit higher crystallinity and dimensional stability but increased brittleness. In addition, composite nanocellulose papers are employed, where the base matrix is modified with polymers or inorganic fillers in order to reduce moisture absorption or intentionally tailor dielectric parameters for specific applications [15].

The primary limitation remains the higher technological complexity of fabrication and limited availability in large-area formats, which restricts their broader deployment in low-cost electronics. Nanocellulose substrates allow controlled adjustment of fibril structure and composition, enabling precise tuning of dielectric properties than conventional paper, which can improve impedance stability in wideband applications.

TABLE I
COMPARISON OF THE ADVANTAGES AND DISADVANTAGES OF DIFFERENT TYPES OF PAPER SUBSTRATES

Substrate Type	Advantages	Disadvantages
Conventional Cellulose Paper	Low cost; wide availability; good flexibility; porous surface enabling strong adhesion of conductive inks	High moisture sensitivity; dielectric instability under humidity variations; possible resonant frequency shifts; higher dielectric losses; ink penetration reducing effective conductivity
Aramid Paper	High mechanical strength; excellent thermal resistance; low hygroscopicity; improved environmental stability	Higher cost; limited availability; weaker adhesion of inks and pastes without additional surface treatment
Nanocellulose Paper	Improved dielectric homogeneity; lower dielectric losses; higher radiation efficiency; reduced ink penetration; biodegradable material	Higher production cost; limited large-area availability; moisture sensitivity in non-composite variants; increased fabrication complexity

III. Ultra-Wideband Antennas

UWB antenna is a broadband antenna designed to operate over an extremely wide frequency range, with a fractional bandwidth exceeding 20% relative to the center frequency or an absolute bandwidth greater than 500 MHz. The effective operational bandwidth is defined using the reflection coefficient parameter S_{11}, which must remain below -10 dB across the entire operating frequency range. In a regulatory context, UWB technology is most commonly defined as a system operating within the 3.1–10.6 GHz frequency band [16], The power spectral density is strictly limited in order to minimize interference with narrowband systems. The absolute bandwidth can be calculated using the following equation [17].

$$BW = f_H - f_L \tag{1}$$

where $f_H$is the upper cutoff frequency and $f_L$is the lower cutoff frequency, both defined at the -10 dB threshold of the S_{11} parameter.

Antenna topology plays a critical role in the design of UWB antennas implemented on paper substrates, as it must compensate for limitations arising from higher dielectric losses and variability in substrate properties. Consequently, the literature predominantly favors planar antenna structures with simple geometries, which provide wide impedance bandwidth and relatively low sensitivity to manufacturing tolerances [16]. Fig. 2. illustrates the typical bandwidth characteristic of a UWB antenna.

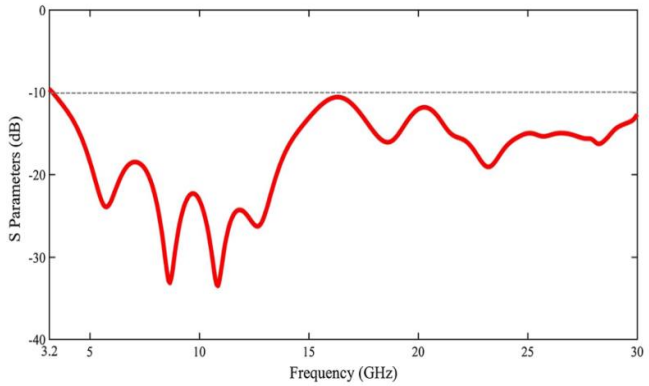


Fig. 2. Reflection coefficient of typical UWB antenna [5]

During the synthesis of a UWB antenna, it is necessary to consider the dielectric properties of the substrate on which the antenna is designed, as the relative permittivity ε_r, dielectric loss tangent $\tan\delta$, substrate thickness h, and their frequency dependence directly influence the effective permittivity of the structure, input impedance, operational bandwidth, and radiation efficiency.

For the design of a simple patch UWB antenna, the antenna dimensions must be calculated based on the dielectric properties of the substrate. The width of the antenna can be determined using the following equation [18].

$$W = \frac{c}{2f\sqrt{\frac{\varepsilon_r + 1}{2}}} \tag{2}$$

where c is the speed of light in free space, $\varepsilon_r$represents the relative permittivity of the substrate, and $f$is the desired center frequency.

The effective permittivity of planar structures can be approximated using the following relation [18].

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1+12\frac{h}{W}\right)^{-\frac{1}{2}} \tag{3}$$

where h is the substrate thickness.

The antenna length can be determined using the following expression, derived from established empirical formulations commonly applied in microstrip antenna design [18].

$$L = \frac{c}{2f\sqrt{\varepsilon_{eff}}} - 0{,}824h\left(\frac{(\varepsilon_{eff} + 0.3)(\frac{W}{h} + 0{,}264)}{(\varepsilon_{eff} - 0{,}258)(\frac{W}{h} + 0{,}8)}\right) \tag{4}$$

When paper-based substrates are employed, their hygroscopic nature and the potential variability of material parameters with ambient humidity must also be considered, as these effects may introduce additional frequency shifts and reduce the reproducibility of measured results.

A. Coplanar waveguide antennas

Coplanar waveguide (CPW) antennas represent one of the most suitable topologies for paper-based substrates. Their primary advantage lies in the realization of both the feed line and ground planes within a single conductive layer, thereby eliminating the need for precise alignment of multiple layers. The CPW topology is less sensitive to substrate thickness and dielectric inhomogeneity, which is particularly advantageous when cellulose-based papers are used. In UWB applications, CPW antennas are frequently combined with monopole or slot radiators, enabling a wide impedance bandwidth while maintaining a simple fabrication process [16].

A potential drawback of CPW antennas is their increased sensitivity to surface roughness and inhomogeneity of the conductive layer, which may result in higher losses [19], [20].

Fig. 3. Design of the manufactured CPW antenna [20]

A notable example of a CPW fed UWB antenna implemented on a paper substrate is a compact circular patch design fabricated by inkjet printing of conductive silver nanoparticle ink on photo paper. The radiating structure consists of a slotted circular patch combined with a reduced, symmetrically slotted ground plane to achieve broadband performance.

The fabricated antenna is illustrated in Fig. 3, where the compact radiator and modified ground structure can be clearly observed. The measured reflection coefficient remains below -10 dB over an extremely wide frequency range spanning approximately 1.66 GHz to 56.1 GHz, corresponding to a fractional bandwidth of about 188.5 %, thereby covering not only traditional UWB allocations but also multiple ISM, WLAN, WiMAX and emerging communication bands. The antenna exhibits a high bandwidth ratio of approximately 33.81:1 and a bandwidth dimension ratio (BDR) among the highest reported for flexible super-wideband antennas of similar class [20].

Furthermore, the realized gain remains substantial across the operating band, and the radiation patterns are relatively stable despite the mechanical flexibility of the paper substrate, which underscores the potential of such printed CPW configurations for robust and low-cost UWB applications [20].

IV. Future work

Future research should address the improvement of moisture resistance in paper-based substrates, as water absorption significantly affects dielectric stability, resonant frequency position, and overall antenna performance. Effective strategies to limit moisture penetration must be developed without introducing materials that would compromise environmental compatibility.

Another critical challenge is the enhancement of material homogeneity. Variations in fiber distribution, porosity, and local density lead to spatial fluctuations in effective permittivity and dielectric losses, reducing reproducibility and long-term stability. Advanced characterization and controlled manufacturing approaches are required to achieve consistent electromagnetic properties.

Overall, the long-term stability of paper substrates under varying environmental conditions remains a central issue. Any modification aimed at improving moisture resistance and dielectric consistency must preserve the primary advantage of paper-based materials—biodegradability. Maintaining this balance between environmental sustainability and electromagnetic reliability represents a key objective for future development.

Currently, experimental work is being conducted on biodegradable paper substrates by fabricating printed UWB antennas based on original design proposals. These experiments aim to evaluate the electromagnetic performance, stability, and practical feasibility of fully biodegradable antenna systems under realistic operating conditions.

References

[1] H. P. Phan, T.-P. Vuong, P. Benech, P. X. P. Borel, and A. Delattre, 'Low-cost wideband antenna on paper substrate', in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, Paris, France: IEEE, Mar. 2017, pp. 2168–2171. doi: 10.23919/EuCAP.2017.7928575.

[2] S. M. Khan, J. M. Nassar, and M. M. Hussain, 'Paper as a Substrate and an Active Material in Paper Electronics', *ACS Appl. Electron. Mater.*, vol. 3, no. 1, pp. 30–52, Jan. 2021, doi: 10.1021/acsaelm.0c00484.

[3] S. Gomez-Gijon, I. Ortiz-Gómez, and A. Rivadeneyra, 'Paper-Based Electronics: Toward Sustainable Electronics', *Adv. Sustain. Syst.*, vol. 9, no. 1, p. 2400486, Jan. 2025, doi: 10.1002/adsu.202400486.

[4] X. Zhou, T. Leng, K. Pan, M. A. Abdalla, and Z. Hu, 'Graphene Printed Flexible and Conformal Array Antenna on Paper Substrate for 5.8GHz Wireless Communications', in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, Copenhagen, Denmark: IEEE, Mar. 2020, pp. 1–4. doi: 10.23919/EuCAP48036.2020.9135742.

[5] T. K. Saha, T. N. Knaus, A. Khosla, and P. K. Sekhar, 'A CPW-fed flexible UWB antenna for IoT applications', *Microsyst. Technol.*, vol. 28, no. 1, pp. 5–11, Jan. 2022, doi: 10.1007/s00542-018-4260-0.

[6] P. Lukacs, T. Lenger, I. Vehec, M. Jusko, and F. Mahdal, 'Homogeneity Analysis of Ecological Paper-Based Substrates' Dielectric Properties', in *2025 International Spring Seminar on Electronics Technology (ISSE)*, Budapest, Hungary: IEEE, May 2025, pp. 1–6. doi: 10.1109/ISSE65583.2025.11121000.

[7] V. Mandrić Radivojević, S. Rupčić, M. Srnović, and G. Benšić, 'Measuring the Dielectric Constant of Paper Using a Parallel Plate Capacitor', *Int. J. Electr. Comput. Eng. Syst.*, vol. 9, no. 1, pp. 1–10, 2018, doi: 10.32985/ijeces.9.1.1.

[8] S. Y. Kim, Y. B. Seo, and J. S. Han, 'Effect of Calendering on the Properties of Paper Containing Flexible Calcium Carbonate with a Cellulose Nanofibril Core', *ACS Omega*, vol. 7, no. 39, pp. 35305–35315, Oct. 2022, doi: 10.1021/acsomega.2c04967.

[9] Y. Zhao, W. Dang, Z. Lu, L. Wang, L. Si, and M. Zhang, 'A novel mica-based composite with hybrid aramid fibers for electrical insulating applications: largely improved mechanical properties and moisture resistance', *Polym. Int.*, vol. 67, no. 2, pp. 204–211, Feb. 2018, doi: 10.1002/pi.5498.

[10] P. Lukacs, A. Pietrikova, I. Vehec, and T. Lenger, 'Long-Term Stability Analysis of the Paper-Based Nomex 410 Substrate's Dielectric Properties', in *2024 International Conference on Diagnostics in Electrical Engineering (Diagnostika)*, Pilsen, Czech Republic: IEEE, Sep. 2024, pp. 1–4. doi: 10.1109/Diagnostika61830.2024.10693901.

[11] F. Hoeng, A. Denneulin, and J. Bras, 'Use of nanocellulose in printed electronics: a review', *Nanoscale*, vol. 8, no. 27, pp. 13131–13154, 2016, doi: 10.1039/C6NR03054H.

[12] A. Li *et al.*, 'Overview of nanocellulose as additives in paper processing and paper products', *Nanotechnol. Rev.*, vol. 10, no. 1, pp. 264–281, May 2021, doi: 10.1515/ntrev-2021-0023.

[13] M. Laudon and B. Romanowicz, Eds, 'TechConnect briefs 2017. Vol. 1: Advanced materials', Danville, CA, U.S.A: TechConnect, 2017.

[14] S. Sami *et al.*, 'Strategies for Enhancing the Dielectric Constant of Organic Materials', *J. Phys. Chem. C*, vol. 126, no. 45, pp. 19462–19469, Nov. 2022, doi: 10.1021/acs.jpcc.2c05682.

[15] M. Wawrzyniak, A. Denneulin, T. P. Vuong, and J. Bras, 'Nanocellulose-based materials and composites for electromagnetism and radio frequencies applications', in *Nanocellulose Based Composites for Electronics*, Elsevier, 2021, pp. 101–124. doi: 10.1016/B978-0-12-822350-5.00005-9.

[16] T. Saeidi, I. Ismail, W. P. Wen, A. R. H. Alhawari, and A. Mohammadi, 'Ultra-Wideband Antennas for Wireless Communication Applications', *Int. J. Antennas Propag.*, vol. 2019, pp. 1–25, Apr. 2019, doi: 10.1155/2019/7918765.

[17] C. Saha, J. Y. Siddiqui, and Y. M. M. Antar, *Multifunctional Ultrawideband Antennas: Trends, Techniques, and Applications*, 1st edn. CRC Press, 2019. doi: 10.1201/9781351026543.

[18] P. Lukacs, A. Pietrikova, I. Vehec, and P. Provazek, 'Influence of Various Technologies on the Quality of Ultra-Wideband Antenna on a Polymeric Substrate', *Polymers*, vol. 14, no. 3, p. 507, Jan. 2022, doi: 10.3390/polym14030507.

[19] Q. Awais, H. T. Chattha, M. Jamil, Y. Jin, F. A. Tahir, and M. U. Rehman, 'A Novel Dual Ultrawideband CPW-Fed Printed Antenna for Internet of Things (IoT) Applications', *Wirel. Commun. Mob. Comput.*, vol. 2018, no. 1, p. 2179571, Jan. 2018, doi: 10.1155/2018/2179571.

[20] Md. R. Hasan, M. A. Riheen, P. Sekhar, and T. Karacolak, 'Compact CPW-fed circular patch flexible antenna for super-wideband applications', *IET Microw. Antennas Propag.*, vol. 14, no. 10, pp. 1069–1073, Aug. 2020, doi: 10.1049/iet-map.2020.0155.

State Representation in Domain-Specific Languages: A Review of Semantic Models

[1]*Monika KOHÚTOVÁ (1st year),*
Supervisor: [2]*William STEINGARTNER*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]monika.kohutova@tuke.sk, [2]william.steingartner@tuke.sk

***Abstract*—Domain-Specific Languages (DSLs) are widely used to model and implement software systems within narrowly defined application domains. A central aspect of many DSLs is the notion of program state and the way state is represented, manipulated, and interpreted during execution. Different classes of DSLs adopt distinct state models, ranging from explicit configurations and stacks to finite control states and data-oriented representations. This paper presents a review of existing approaches to state representation in domain-specific languages and their associated semantic descriptions. The survey covers imperative, stack-based, finite-state, and data-oriented DSLs, highlighting similarities and differences in how state is modeled and transformed. Rather than proposing a new semantic framework, the paper organizes existing work according to state-centric design choices and semantic styles. The presented comparison provides a structured perspective on current practices in DSL design and semantics, serving as a foundation for further analysis and discussion.**

***Keywords*—domain-specific languages, program state, formal semantics, state-based systems, semantic modeling**

I. Introduction

Program state is a core notion in the formal description of computation. At an abstract level, a state captures the information required to characterize the current configuration of an executing system (e.g., memory contents, control position, or execution context), and program behavior can be viewed as the evolution of such configurations over time [1], [2].

Different paradigms and language families expose and manipulate state in different ways. Imperative languages typically model state as an explicit mutable store, while declarative settings may encode state changes implicitly. Similarly, stack-based and concatenative languages treat the stack as the primary carrier of state, whereas automaton-based models identify state with control configurations [3]. These choices reflect distinct semantic assumptions about computation and directly affect analyzability.

Domain-Specific Languages [4], [5] provide a particularly suitable setting for studying state representation, because they align language constructs and semantics with domain concepts. In contrast to general-purpose languages, DSLs deliberately restrict expressiveness in favor of domain alignment, which enables more structured state models and clearer semantic descriptions [4]. Consequently, DSLs employ a variety of state models, including configurations, stacks, finite control states, and data-oriented representations [6], [7]. Formal semantics [1] helps make these design choices precise and supports systematic reasoning about correctness.

This paper reviews state representation in domain-specific languages and the corresponding semantic styles, without proposing a new semantic framework. Section II introduces the conceptual background used in the paper. Section III surveys representative DSL classes organized by their dominant state models and provides a comparative summary. Section IV synthesizes recurring patterns and trade-offs across these classes and Section V concludes the paper.

II. Program State as a Conceptual Lens

To compare domain-specific languages, it is necessary to clarify the notion of *program state* and its interpretation across semantic frameworks. Rather than introducing a new definition, this section adopts a conceptual perspective on program state consistent with established approaches in programming language semantics.

A. Program State and Its Representation

In programming language theory, the state of a program is commonly understood as a mathematical abstraction [1], [2] capturing all information required to describe the current configuration of program execution. Depending on the language paradigm and semantic model, this information may include variable bindings, memory contents, control locations, stacks, or external resources.

We denote the state space by $\mathbf{State}$. A widely used formal abstraction represents program execution as state transformation. Following standard semantic practice, a program can be viewed as a function acting on states:

$$\delta : \mathbf{State} \to \mathbf{State}, \tag{1}$$

where δ represents a state transition induced by program execution.

While this abstraction provides a clean formal basis, it assumes total and deterministic state transformations. In practical settings, program execution may exhibit partiality due to runtime errors, non-termination, or undefined behavior. Such phenomena can be captured by extending the model to partial functions or transition relations over $\mathbf{State}$.

This abstraction, employed in various forms in operational, denotational, and coalgebraic semantics [8], [9], emphasizes that computation can be understood as the evolution of state over time.

The structure of the state space $\mathbf{State}$ depends on the language. In practice, program states may be represented in different ways:

- a mapping from variables to values,
- a stack-based configuration,
- a finite control state of an automaton,
- a database instance or structured data store.

Although these representations differ significantly, they all conform to the same abstract view of computation as state transformation.

B. State and Semantic Interpretation

The interpretation of program state is closely tied to the chosen semantic framework. Different formal approaches provide distinct yet complementary views on program behavior. We consider a program P as a finite sequential composition of statements

$$P = S_1;\, S_2;\, \cdots ;\, S_n, \tag{2}$$

where each S_i denotes a statement of the language.

In *operational semantics*, program execution is described by explicit transitions between states, typically formalized as a transition relation over configurations [1]:

$$\langle S, s\rangle \;\rightarrow\; \langle S', s'\rangle, \tag{3}$$

where S and S' denote program fragments and $s, s' \in \mathbf{State}$ represent program states. This approach makes state change explicit and enables stepwise reasoning about program execution.

In *denotational semantics* [2], programs are mapped to mathematical objects that represent their meaning, often as functions over semantic domains. A common state-based interpretation assigns to a program P a semantic function

$$[\![P]\!] : \mathbf{State} \rightarrow \mathbf{State}, \tag{4}$$

emphasizing compositionality and abstraction from intermediate execution steps.

In *axiomatic semantics*, the focus shifts to reasoning about correctness by relating states before and after execution. This perspective is commonly expressed using Hoare-style assertions introduced by Hoare [10]:

$$\{\,\varphi\,\}\; P\; \{\,\psi\,\}, \tag{5}$$

where φ and ψ are predicates over $\mathbf{State}$. Rather than describing how execution proceeds, axiomatic semantics specifies logical conditions that characterize valid state transformations.

These semantic perspectives are frequently employed in the formal description of domain-specific languages. In particular, DSLs benefit from semantic models that make state transformations explicit, supporting reasoning about correctness, safety, termination, predictability, and domain-specific constraints [11].

C. Aspects for Comparing Domain-Specific Languages

Based on the above considerations, the comparison of DSLs in this paper is guided by several conceptual aspects related to state representation and manipulation:

- the structure of the program state,
- the degree to which state is represented explicitly or implicitly,
- the semantic style used to describe state transitions,
- the relationship between state representation and execution model.

These aspects provide a basis for the comparative review in the next section.

D. Illustrative Model of State Transformation

Figure 1 illustrates the abstract view of program execution adopted in this paper. A program is interpreted as an entity that transforms an initial state into a subsequent one, regardless of the concrete representation used by a particular language.

Fig. 1. Abstract model of program execution as state transformation

This abstract model serves as a unifying conceptual reference for the diverse classes of domain-specific languages discussed in the following section.

III. Overview of Domain-Specific Languages and State Models

Domain-Specific Languages are designed to address problems within a narrowly defined application domain, allowing language constructs and semantics to be closely aligned with domain concepts. As a result, DSLs often expose the notion of program state more explicitly and in a more structured manner than general-purpose languages. This section provides a comparative overview of several classes of DSLs, organized according to the way they represent and manipulate program state.

A. Conceptual Classification of Domain-Specific Languages

Before presenting individual DSLs in detail, it is useful to introduce a high-level conceptual classification based on their dominant form of program state representation. This classification uses the notion of state as a unifying abstraction for positioning different language families within a common conceptual space.

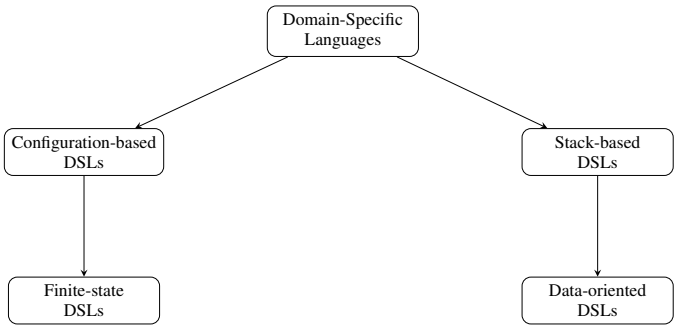


Fig. 2. Conceptual classification of domain-specific languages by dominant state representation

Figure 2 illustrates a simplified taxonomy of domain-specific languages based on their primary state representation. The classification does not aim to be comprehensive, nor does it imply strict boundaries between categories. Instead, it highlights dominant design tendencies that recur across multiple DSLs used in control-oriented and formally specified systems.

TABLE I
COMPARISON OF DOMAIN-SPECIFIC LANGUAGES WITH RESPECT TO STATE REPRESENTATION AND SEMANTIC PERSPECTIVE

DSL Class	Representative Languages	State Representation	Semantic Perspective
Imperative / Mission DSLs	MBL, PLEXIL, Robot DSLs	Explicit configuration (variables, control, environment)	Operational semantics
Stack-Based DSLs	StackLang, Joy, KKJ	Stack as sole state carrier	Denotational / compositional
Finite-State DSLs	FSM-based DSLs, RAM	Finite control states, memory cells	Transition systems
Data-Oriented DSLs	SubSQL	Database instance	Declarative semantics

B. Imperative and Mission-Oriented DSLs

Imperative DSLs model computation as transformations over an explicitly defined state space, typically consisting of evaluation memory, internal control context, and system configuration. Execution proceeds through statements that modify this structured representation. Representative examples include PLEXIL and MBL, used in mission-critical and autonomous control environments.

These languages are common in control- and mission-oriented domains requiring predictable state evolution. Mission-oriented DSLs are designed for autonomous systems operating without direct human intervention, for example in air-traffic control or space missions (notably NASA applications), where programs govern device behavior under strict safety and reliability constraints [12], [13].

State transitions are defined operationally, with each statement corresponding to a deterministic transformation of the current configuration. This supports reasoning about execution order and verification of safety and liveness properties, often within a property-driven design framework [13].

Although imperative DSLs emphasize explicit state manipulation, some domain-specific languages operate in a restricted or stateless manner; even in these cases, program behavior can still be systematically interpreted within an abstract state-based semantic framework [14].

C. Stack-Based and Concatenative DSLs

Stack-based and concatenative DSLs represent program state primarily as a stack structure. Computation is expressed through the composition of operations that manipulate the stack, such as push, pop, and stack reordering primitives. Representative examples include StackLang and the concatenative languages Joy and KKJ [11], [15], [16].

In this class of languages, the stack itself constitutes the complete program state. The absence of named variables leads to a uniform and compact semantic model, where program execution corresponds to successive transformations of the stack. This representation supports compositional reasoning and has been shown to be well suited for formal semantic descriptions and language generation frameworks [11].

D. Finite-State and Control-Oriented DSLs

Finite-state and control-oriented DSLs model computation in terms of transitions between a finite set of control states [17].

In this setting, the program state is identified with the current control state of an automaton, possibly enriched with additional data components. State transitions are triggered by events or conditions and are typically described declaratively. Related models include Random Access Machines (RAM), which can be interpreted as DSLs focused on memory-state manipulation through a fixed instruction set [18].

E. Data-Oriented DSLs: The Case of SubSQL

Data-oriented DSLs focus on the manipulation of structured data rather than explicit control flow. The language SubSQL represents a representative example of this class, providing a domain-specific abstraction for database-oriented computations. In SubSQL, the program state corresponds to a database instance, and program execution is expressed as transformations of this state through declarative queries and updates [12].

Unlike imperative DSLs, state changes in SubSQL are described implicitly through high-level operations, emphasizing *what* transformation should be performed rather than *how* it is executed. This approach highlights a different semantic perspective on state, where the underlying execution mechanism is abstracted away in favor of declarative expressiveness.

F. Comparative Summary

Table I summarizes the key characteristics of state representation and semantic perspective across DSL classes. The comparison highlights how DSLs conceptualize program state and express state transformations within their semantic models.

Despite differences in execution models [6], all reviewed DSLs rely on an explicit notion of state, which serves as a unifying semantic concept.

IV. COMPARISON AND SYNTHESIS

The review in Section III shows that domain-specific languages differ in syntax and execution models, yet consistently rely on an explicit notion of state and state transformation. This section summarizes shared patterns and key differences across the surveyed DSL classes, focusing on the conceptual role of state rather than implementation details.

A. Shared Patterns Across DSL Classes

Across imperative, stack-based, finite-state, and data-oriented DSLs, a recurring structural pattern can be identified: languages define (implicitly or explicitly) a set of states and provide constructs that describe how these states evolve during execution. In particular, the following commonalities emerge:

- *State as a first-class semantic notion:* even when state is not explicitly named (e.g., in declarative settings), program meaning depends on an underlying configuration that changes during execution.
- *State transformation as the core of behavior:* DSLs typically emphasize how domain-level actions map to

transitions between configurations (operationally) or to transformations of abstract semantic objects.
- *Domain constraints reflected in state structure:* the choice of state representation is closely related to domain concepts (e.g., control modes, resource levels, stack contents, or database instances).

These shared patterns support a uniform conceptual reading of DSL semantics as state-based interpretation, consistent with the abstract model introduced in Section II.

B. Key Differences and Trade-offs

While most DSL classes employ state-based reasoning, they differ in the way state is represented and exposed to the programmer and in how state changes are described. The following differences are particularly relevant from a semantic perspective:

- *Explicit vs. implicit state:* imperative and finite-state DSLs often make state explicit through configurations or control states, whereas data-oriented DSLs may encode state changes through declarative operations.
- *Continuous vs. discrete state spaces:* stack- and configuration-based DSLs typically operate over richer state spaces, while finite-state DSLs restrict behavior to discrete control states, enabling stronger analysis at the cost of expressiveness.
- *Operational vs. denotational emphasis:* some DSLs are described through stepwise transitions, while others use compositional semantic descriptions that treat programs as transformations over abstract domains.

These trade-offs reflect design priorities: analyzability and predictability often motivate restricted state models, whereas compositionality motivates richer state representations.

C. Summary

In summary, DSLs provide structurally comparable approaches to modeling program behavior through state. The reviewed literature indicates that state representation is not merely an implementation concern but a central semantic design choice, often shaped by required system properties in a property-driven manner [13]. By organizing DSLs according to their dominant state model, the review clarifies recurring design tendencies and shows how semantic interpretation follows from the underlying notion of state.

V. Conclusion

This paper presented a concise review of state representation in domain-specific languages and related semantic descriptions. Rather than focusing on syntactic variability, the discussion emphasized state as the central semantic interface between a DSL and its execution model.

The proposed organization clarifies the literature by distinguishing DSL classes according to the form of state they expose (or abstract away) and the semantic style used to describe state change.

The analysis also reveals several open challenges. First, the commonly used assumption of total and deterministic state transformations does not fully capture partial or undefined behavior observed in real systems. Second, integrating stateless or purely functional DSLs into a unified state-based semantic framework remains conceptually non-trivial. Third, the relationship between state representation and formal verification techniques is not yet systematically explored across DSL classes.

Future research will focus on extending state-based models to incorporate partiality and non-determinism, improving compositional reasoning across heterogeneous DSLs, and developing verification-oriented semantic frameworks tailored to domain-specific constraints. These directions aim to strengthen the role of formal semantics in the design, analysis, and validation of domain-specific languages.

Acknowledgment

This work was supported by KEGA project 028TUKE-4/2026 – "Innovative approaches to teaching and researching formal methods in the context of modern software engineering", granted by the Cultural and Education Grant Agency of the Slovak Ministry of Education.

References

[1] H. R. Nielson and F. Nielson, *Semantics with Applications: An Appetizer*, ser. Undergraduate Topics in Computer Science. Springer, 1999.
[2] J. E. Stoy, *Denotational Semantics: The Scott–Strachey Approach to Programming Language Theory*, ser. Series in Computer Science. Cambridge, MA: MIT Press, 1977.
[3] P. Sestoft, *Programming Language Concepts*. Springer, 2012.
[4] M. Fowler, *Domain Specific Languages*. Addison-Wesley Professional, 2010.
[5] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Computing Surveys*, vol. 37, no. 4, pp. 316–344, 2005.
[6] T. Kosara, S. Bohrab, and M. Mernik, "Domain-specific languages: A systematic mapping study," *Information and Software Technology*, vol. 65, pp. 1–17, 2015.
[7] M. Völter, S. Benz, C. Dietrich, B. Engelmann, M. Helander, L. C. L. Kats, E. Visser, and G. Wachsmuth, *DSL Engineering: Designing, Implementing and Using Domain-Specific Languages*. North Charleston, SC, USA: CreateSpace Independent Publishing Platform, 2013.
[8] W. Steingartner, V. Novitzká, M. Bačíková, and Š. Korečko, "New approach to categorical semantics for procedural languages," *Computing and Informatics*, vol. 36, no. 6, pp. 1385–1414, 2017.
[9] W. Steingartner, V. Novitzká, and W. Schreiner, "Coalgebraic operational semantics for an imperative language," *Computing and Informatics*, vol. 38, no. 5, pp. 1181–1209, 2019.
[10] C. A. R. Hoare, "An axiomatic basis for computer programming," *Communications of the ACM*, vol. 12, no. 10, pp. 576–580, 1969.
[11] J. Mihelič, W. Steingartner, and V. Novitzká, "A denotational semantics of a concatenative/compositional programming language," *Acta Polytechnica Hungarica*, vol. 18, no. 4, pp. 231–250, 2021. [Online]. Available: https://acta.uni-obuda.hu/Mihelic_Steingartner_Novitzka_111.pdf
[12] W. Schreiner and W. Steingartner, "Semantics-based rapid prototyping of a machine controller language," in *2024 IEEE 17th International Scientific Conference on Informatics*, V. Novitzká and A. Szakál, Eds. IEEE, 2024, pp. 337–342. [Online]. Available: https://doi.org/10.1109/Informatics62280.2024.10900792
[13] M. Brambilla, C. Pinciroli, M. Birattari, and M. Dorigo, "Property-driven design for swarm robotics," in *Swarm Intelligence*, ser. Lecture Notes in Computer Science, vol. 7461. Springer, 2012, pp. 139–146.
[14] B. Jacobs, *Introduction to Coalgebra: Towards Mathematics of States and Observation*, ser. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2016, vol. 59.
[15] W. Schreiner and W. Steingartner, "The SLANG semantics-based language-generator: Tutorial and reference manual (version 1.0)," Research Institute for Symbolic Computation, Johannes Kepler University Linz, Technical Report 23-13, 2023. [Online]. Available: https://doi.org/10.35011/risc.23-13
[16] M. von Thun, "Joy — a purely functional language," *Department of Computer Science, La Trobe University*, 2001. [Online]. Available: https://www.kevinalbrecht.com/code/joy-mirror/j05cmp.html
[17] E. Fülöp and N. Pataki, "A DSL for resource checking using finite state automaton driven symbolic execution," *Open Computer Science*, vol. 11, no. 1, pp. 107–115, 2021.
[18] W. Steingartner, S. Novotný, and P. Gnip, *Formal Languages and Automata: From Theory to Practice*. Košice, Slovakia: Technical University of Košice, 2026, (*in Slovak*).

Automated Detection of Aurora Borealis in Mid-Latitudes Using Public Camera Networks

[1]*Martin NGUYEN (4th year),*
Supervisor: [2]Ján GENČI, Consultant: [3]Pavol BOBÍK

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[3]Department of Cosmic Physics, Institute of Experimental Physics SAS Kosice, Slovak Republic

[1]martin.nguyen@student.tuke.sk, [2]jan.genci@tuke.sk, [3]bobik@saske.sk

***Abstract*—This paper presents the development of an automated detection and alerting system for Aurora Borealis observations in the mid-latitude region of Slovakia by public cameras. Due to the rarity of these events at lower geomagnetic latitudes and the limitations of manual monitoring, researchers often miss valuable observation opportunities. The proposed system utilizes public meteorological camera feeds and the OpenAI CLIP algorithm to perform semantic image analysis on standard web servers. Unlike traditional color-based methods, this solution uses zero-shot learning to identify auroral features out of the box. Evaluation on the OATH benchmark dataset demonstrated a precision of 90.49% and an F1-score of 0.90, while real-world testing across a camera network achieved a 98% success rate in detecting events, including the major geomagnetic storms of 2024 and 2026. The system provides a cost-effective and real-time alternative to manual surveillance, with the possibility to automatically send reports.**

***Keywords*—Aurora Borealis, Automated Detection, CLIP Model, Public Camera Networks**

I. Introduction

Since 2023, Slovakia has seen a significant increase in northern lights sightings, even at mid-latitudes (approx. $48° - 49°$N). These events have become very popular on social media and in news reports. While amateur photographers have successfully captured these occurrences, professional researchers at the Institute of Experimental Physics (IEP SAS) still find it difficult to track them in real time. Currently, detection often relies on manually monitoring the Kp-index. However, this method is insufficient because high geomagnetic activity does not guarantee local visibility due to variables like cloud cover and light pollution.

Past efforts to address this gap included using small Raspberry Pi cameras, but these were limited by their reliance on constant human supervision. The manual review was necessary and often resulted in missed opportunities during short-duration events. This research introduces a fully automated, low-resource detection system. It is designed to ensure that rare events, such as the occurrence in January 2026, are captured, recorded for the scientific community, and reported to researchers in real time.

II. CLIP Model (Contrastive Language-Image Pre-training)

OpenAI's CLIP architecture [1] moves away from traditional supervised learning computer vision, which requires training of the model and is usually limited to a fixed list of categories. Traditional models struggle to identify rare events like the Aurora Borealis if those events were not included in their original training data. CLIP solves this by using natural language to understand visual concepts through a dual-encoder system. It consists of a vision encoder for images and a text encoder for descriptions that work together in a shared mathematical space.

The CLIP was trained on the WIT dataset [2], which contains 400 million image-text pairs collected from the internet. During training, the model learns to match correct images with their corresponding text descriptions. This process allows the model to perform zero-shot transfer, meaning it can recognize images it has never specifically been trained on by using simple text prompts. For atmospheric research, this is a major advantage because researchers can immediately deploy the system using specific descriptions. Instead of complex programming, the model can identify terms like "green aurora in the dark sky" or "aurora borealis over the northern horizon," because it understands the actual meaning of the words rather than just looking at pixel patterns.

Aside from CLIP, other deep learning models have shown high success in classifying auroras, though they typically require large, hand-labeled datasets and constant supervision. For example, researchers using the ConvNeXt model on high-resolution images from the THEMIS network achieved an impressive 99.4% accuracy when identifying specific shapes like arcs or diffuse auroras [3]. However, standard models like ResNet-50 and EfficientNet can struggle with the fast-moving nature of auroras or with "no-aurora" images that contain complex city lights and clouds [4]. While unsupervised methods like SimCLR can group images into clusters without manual labels, they still require significant human effort to interpret the results and determine what each cluster represents [5]. These supervised and unsupervised systems are highly effective in stable, high-latitude environments, but they lack the "out-of-the-box" flexibility needed for real-time monitoring in mid-latitude regions like Slovakia.

III. System architecture

Our system uses a Python FastAPI backend and an Angular frontend hosted on a cloud server. Instead of using old methods that only look at colors, it uses the OpenAI CLIP model ($openai/clip - vit - large - patch14$) to understand

the actual meaning of an image. This allows our system to recognize an aurora "out of the box", without additional training and without needing thousands of photos of Aurora from Slovakia. Every five minutes, a script captures images from six Meteopress cameras facing North and analyzes them using the CLIP model. The algorithm compares each image against "positive" descriptions (like "green aurora in the dark sky") and "negative" descriptions (like "clear night sky"). If the model is at least 65% sure ($C_{aurora} \geq 0.65$) that an aurora is present, it automatically flags a positive detection.

By using FastAPI's background tasks, the system is highly efficient and is capable of finishing the entire detection in less than 10 seconds on regular CPU, without requiring expensive hardware. The system was successful during the January 2026 aurora event by sending an automated alert while researchers were away. The event was captured and recorded in real time.

Our system uses a specialized sub-network of Meteopress meteorological cameras, specifically choosing six locations — Svit, Liptovská Osada, Stanča, Rybany, Reichersberg, and Lúčky. They offer free real-time access to clear and North-facing views. To keep the backend lightweight, we developed a module that connects directly to real-time HLS (HTTP Live Streaming) video feeds. Instead of downloading heavy video files, the service extracts a single high-resolution frame from the stream buffer every five minutes.

This architecture allows the system to remain highly efficient and run on affordable hosting without expensive hardware. By capturing frames at 5-minute intervals, the system is fast enough to detect short-lived "substorms" and still provide accurate AI classification. An additional feature is the historical fetch feature, which allows us to retroactively pull frames from the past to validate our results.

IV. Results and Evaluation

To objectively assess the performance of our CLIP-based detection system, we made a validation experiment using a standardized dataset. While the system is designed for real-time detection in Slovakia, evaluating it against a known benchmark ensures the reliability of the reported results. We also tested the system on real camera images by pulling saved frames from past aurora sightings.

A. Evaluation on the OATH Dataset

We utilized the OATH (Oslo Auroral THEMIS) dataset [6] of classified auroral images. This dataset originally categorizes images into seven distinct classes: arc, diffuse, discrete, noaurora, clear, moon, and cloudy. Since our system operates as a binary detector (classifying the presence or absence of aurora), we mapped these original OATH labels into two binary categories for this evaluation:

- Aurora (Positive): Includes images labeled as arc, diffuse, and discrete.
- No Aurora (Negative): Includes images labeled as noaurora, clear, moon, and cloudy.

The OATH dataset utilizes all-sky (fisheye) imagery, as an example shown in the figure 1.

We applied our CLIP-based detector to this dataset using a confidence threshold of $\tau = 0.80$. This means any image where the model predicted an aurora probability score of 0.80 or higher was classified as a positive detection. This higher threshold was chosen to prioritize higher confidence in alerts.

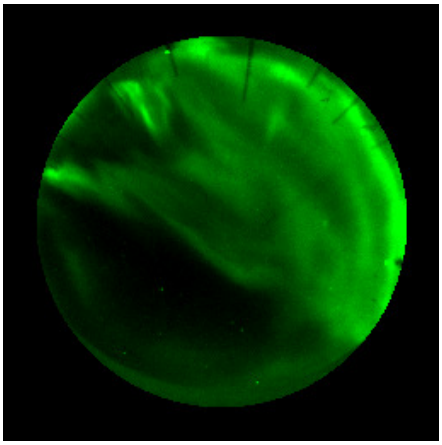


Fig. 1. A sample frame from the OATH benchmark dataset [6] showing a standard all-sky (fisheye) projection dominated by green emission.

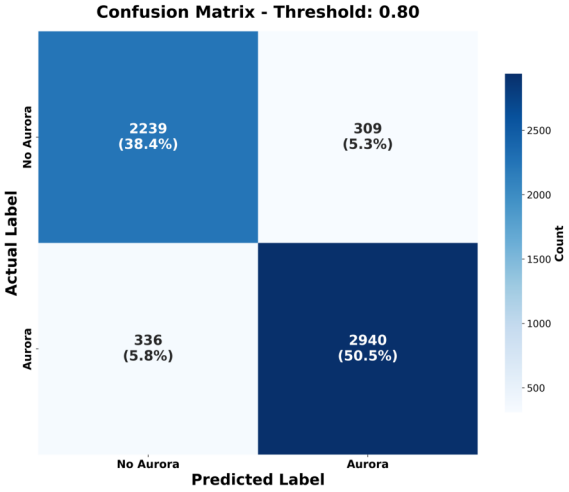


Fig. 2. Confusion Matrix ($\tau = 0.80$), classification of OATH Database [6]

The evaluation was performed on a total of 5,824 images. The performance distribution is visualized in the Confusion Matrix, shown in the figure 2.

The system successfully identified 2,940 auroral displays (TP) and correctly rejected 2,239 non-auroral skies (TN). In terms of classification errors, the model generated 309 false positives (FP), and missed 336 auroral events (FN). We calculated standard performance metrics to assess the system's reliability.

The Precision was 90.49%, which is very important for our automatic alert system. It means that when the system sends an email, there is a 90% chance that the aurora is actually there. This prevents researchers from getting too many "false alarms", which can be annoying and a waste of time.

At the same time, the Recall was 89.74%. This shows that the system found almost all of the aurora events in the data. The Specificity was 87.87% and proved that the system is good at telling the difference between a real aurora and confusing things like moonlight or clouds. The last measure is an F1-Score of 0.9011. The system proves it is a reliable tool as it does not sacrifice detection capability to achieve high precision.

B. Evaluation on web cameras

We selected high-quality, colorful imagery from six locations — Svit, Liptovská Osada, Stanča, Rybany, Reichersberg,

Fig. 3. A successful positive detection from the Rybany camera on January 19, 2026.

and Lúčky. Our evaluation focused on a set of 10 distinct historically confirmed auroral events on 430 frames where auroral activity was confirmed and 842 frames with no auroral activity. In this real-world test, we confirmed that the CLIP algorithm works well with the Slovak camera network and can accurately recognize the aurora in different locations. An example of a successful positive detection from the Rybany camera on January 19, 2026 is shown in the figure 3.

The results of our testing are following:

- The algorithm achieved a 98% success rate in detecting auroras captured on camera feeds. The remaining 2% (False Negatives) occurred when the aurora was extremely faint. In these cases, the Aurora light was so dim that it was barely visible even to the human eye. The system's confidence score was still 30–40% on these events, which is below our threshold for sending an alert.
- In a control group test (20 days with zero auroral activity), the tool correctly identified that there were no auroras 100% of the time.
- A threshold of 0.65 proved to be the ideal. At this value, the tool reliably ignored common light pollution (e.g. Svit camera with a strong red light from a traffic signal was not flagged as Aurora).

The tool detects activity regardless of whether the aurora is green, red, pink, or orange. This was confirmed during the G4 Storm (January 19, 2026), when cameras in Rybany, Lúčky, and Reichersberg captured a spectrum ranging from green to red. The tool identified the Aurora on all of the cameras successfully.

Our testing showed that a 30-minute sampling interval is insufficient. Auroras are highly dynamic. For instance, on October 10, 2024, the phenomenon appeared for only an hour. Switching to a 5-frame burst analysis allowed us to capture short and intense flashes.

The input quality is also important. Based on long-term observation (2023–2026), we evaluated the performance of individual locations:

- Liptovská Osada: This is currently the top-performing location. It is a new, high-quality camera pointed at the stars combined with minimal light pollution. Historical data before 2025 are missing.
- Stanča: It successfully detected even extremely faint pillars, but is susceptible to cloud cover.
- Svit: A problematic location (frequent fog, water droplets on the lens, light noise). However, it provided valuable data during strong storms (like the G5 in May 2024).
- Reichersberg and Lúčky: Reichersberg serves primarily as a backup (lower quality). Lúčky shows promise for future measurements, as historical data before 2025 are missing

V. Conclusion

This research introduces a new way to automatically detect the northern lights in Slovakia using the CLIP model. Unlike older methods that only look at colors, our system understands the actual meaning of an image. This makes it efficient enough to run on cheap web servers without expensive hardware.

Our tests showed the system is very reliable. After analyzing over 400 aurora images from 2023 to 2026, we found that a 65% confidence threshold works best for our region. At this level, the system correctly identified 98% of real aurora events, including the major storm in January 2026. It also correctly ignores distractions like red streetlights. We also found that checking images every five minutes is essential for catching short-lived substorms that only last few minutes.

The success of this project proves that we can use existing weather cameras to monitor space weather automatically. Even with different camera types and city lights, the CLIP model successfully identified auroras with high confidence. Our system ensures that rare events are never missed and provides valuable data for scientists without requiring anyone to manually watch the screens.

The Key Historical Milestones were spotted on:

- May 10, 2024 (Kp-max=9): A historic moment where detection was 100% successful across all functional cameras, despite facing local weather challenges.
- January 19, 2026 (Kp-max=8.7): Auroras were detected on all monitored cameras simultaneously. This confirmed the tool's ability to process data from diverse sources in real-time.

References

[1] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," 2021. [Online]. Available: https://arxiv.org/abs/2103.00020

[2] K. Srinivasan, K. Raman, J. Chen, M. Bendersky, and M. Najork, "Wit: Wikipedia-based image text dataset for multimodal multilingual machine learning," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 2443–2449. [Online]. Available: https://doi.org/10.1145/3404835.3463257

[3] S. ZhiYuan, Y. ZhongHua, L. Jian, X. LinLi, X. Yan, Z. BinZheng, G. RuiLong, Y. Yuan, and W. Yong, "Statistical study of auroral variability under different solar wind conditions based on classification using deep learning techniques," *Earth and Planetary Physics*, vol. 9, no. 6, p. 1163–1170, 2025. [Online]. Available: http://dx.doi.org/10.26464/epp2025075

[4] J. Chen, P. Zhou, Y. Hua, D. Chong, M. Cao, Y. Li, W. Chen, B. Zhu, J. Liang, and Z. Yuan, "Climateiqa: A new dataset and benchmark to advance vision-language models in meteorology anomalies analysis," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, ser. KDD '25. ACM, Aug. 2025, p. 5322–5333. [Online]. Available: http://dx.doi.org/10.1145/3711896.3737406

[5] N. Partamies, B. Dol, V. Teissier, L. Juusola, M. Syrjäsuo, and H. Mulders, "Auroral breakup detection in all-sky images by unsupervised learning," *Annales Geophysicae*, vol. 42, no. 1, p. 103–115, Apr. 2024. [Online]. Available: http://dx.doi.org/10.5194/angeo-42-103-2024

[6] L. B. N. Clausen and H. Nickisch, "Automatic classification of auroral images from the oslo auroral themis (oath) data set using machine learning," *Journal of Geophysical Research: Space Physics*, vol. 123, 2018.

Analog Frontend for Measuring Bioimpedance

[1]*Miroslav JUSKO (2nd year)*
Supervisor: [2]Peter LUKÁCS

[1,2]Dept. of Technologies in electronics, FEI, Technical University of Košice, Slovak Republic

[1]miroslav.jusko@tuke.sk, [2]peter.lukacs@tuke.sk

***Abstract*—This paper presents the design and validation of an analog front end (AFE) for tetrapolar bioimpedance measurement based on the AD5933 impedance converter. Since the AD5933 is not directly suitable for safe human measurements, a dedicated front end was developed, including a voltage controlled current source, differential voltage sensing stage, and calibration switching network. The system provides a controlled, DC free excitation current compliant with IEC 60601-1 requirements and enables accurate four-wire impedance measurement. Experimental results obtained from precision resistors, an RC skin equivalent phantom, and in vivo testing show magnitude error below 2% (for loads above 100 Ω) and confirm the ability to detect pulse wave related impedance variations.**

***Keywords*—Analog frontend, Bioimpedance, Tetrapolar measurement, Pulse wave.**

I. Introduction

Bioimpedance measurement is a noninvasive and cost effective technique widely used for assessing physiological and tissue-related parameters [1]. This paper presents the design and validation of an analog front end (AFE) for tetrapolar bioimpedance measurement based on the AD5933 impedance converter. Although the AD5933 integrates signal generation and impedance calculation, its native configuration is not directly suitable for safe human bioimpedance measurements. To address these limitations, a dedicated analog front end was developed, consisting of a voltage controlled current source, differential voltage sensing stage, and calibration switching network.

The proposed system converts voltage excitation into a controlled, DC free current compliant with IEC 60601-1 safety requirements and enables accurate four-wire impedance measurements across a broad frequency range. Calibration procedures are implemented to compensate for gain and phase errors introduced by the analog signal chain.

The performance of the system was evaluated using precision resistive loads, a skin equivalent RC phantom, and in vivo measurements. Experimental results demonstrate magnitude error below 2% for resistances above 100 Ω and confirm the ability to detect small impedance variations associated with the arterial pulse wave. The proposed design provides a compact and low complexity solution suitable for research oriented bioimpedance applications.

II. Theoretical background

Bioimpedance measurement systems are widely used for body composition analysis, tissue characterization, and cardiovascular monitoring. Contemporary bioimpedance instruments typically employ a tetrapolar electrode configuration to reduce the influence of electrode-skin interface impedance and improve measurement accuracy.

From an instrumentation perspective, such systems generally consist of a sinusoidal signal generator, followed by a current injection stage. The voltage response of the tissue is then amplified using a differential sensing circuit and converted into digital form by an analog-to-digital converter (ADC). Magnitude and phase of the impedance are subsequently obtained using digital signal processing techniques.

Although this architecture is well established and provides flexibility in frequency selection and signal analysis, its performance strongly depends on the design of the analog front end. In particular, accurate bioimpedance measurement requires stable and controlled current injection and high common-mode rejection ratio. Furthermore, when measurements are performed on human subjects, the excitation stage must comply with medical safety standards.

Another critical design parameter is the operating frequency range. Biological tissues exhibit frequency dependent electrical behavior characterized by dispersion regions. In practical biomedical applications, the β-dispersion region (typically from several kilohertz to several megahertz) provides the most relevant diagnostic information, as it reflects both extracellular and intracellular conduction mechanisms [2]. Bioimpedance systems intended for human use must satisfy IEC 60601-1 safety requirements. The injected alternating current must remain within safe limits typically below 100 μA at frequencies above 10 kHz and direct current components must be minimized to prevent electrode polarization and electrochemical reactions at the skin–electrode interface [3]. These constraints impose strict requirements on current control and DC suppression within the analog front end.

The proposed analog front end builds upon the architecture previously introduced in [4], where the basic concept of a current controlled excitation stage and phase sensitive detection was presented. In this work, additional experimental results and extended validation are presented.

III. Bioimpedance measurement analog frontend

Based on the design requirements discussed in Section II, an analog front end (AFE) for bioimpedance measurement was developed to enable safe, accurate, and flexible impedance acquisition over a wide frequency range. The proposed AFE extends the capabilities of the AD5933 impedance converter, which in its native configuration is not fully suitable for biological impedance measurements due to voltage excitation, two-wire operation, and limited current control.

The overall system architecture is shown in Fig. 1. The AFE consists of three main functional blocks: a voltage controlled current source, a differential voltage sensing stage, and a digitally controlled switching network used for calibration and measurement configuration.

A. Signal Generation and Acquisition

The core signal generation and digitization is performed by the AD5933 impedance converter. The device integrates a programmable sinusoidal excitation source, a 12-bit ADC, and an internal discrete Fourier transform (DFT) engine that computes the real and imaginary components of the measured impedance.

In the proposed system, the AD5933 is used as a signal generator and synchronous demodulator rather than as a complete impedance measurement solution. An external microcontroller configures the excitation parameters, controls the AFE, performs calibration, and handles data communication with a host computer via an I²C interface.

Because the AD5933 provides voltage excitation with a frequency dependent DC bias, its output cannot be directly connected to electrodes in contact with the human body. Therefore, a dedicated analog front end was designed to convert the excitation voltage into a controlled, DC free alternating current suitable for tetrapolar bioimpedance measurements.

B. Analog Frontend Design

To ensure compliance with safety requirements and minimize the influence of electrode-skin interface impedance, the excitation signal is applied through a voltage controlled current source. The current source is based on a mirrored Howland topology, which allows precise control of the injected current while maintaining high output impedance.

The excitation voltage generated by the AD5933 has a configurable amplitude ranging from 198 mV to 1.98 V and includes a DC bias dependent on the selected output range. To suppress this DC component, a low-pass filter is applied at the generator output. The filtered signal is then used as a control voltage for the current source.

The output current amplitude I_{out} is primarily determined by a reference resistor R_0, such that:

$$I_{out} = \frac{V_{in}}{R_0} \tag{1}$$

where V_{in} is the filtered excitation voltage. By appropriate selection of R_0, the injected current is limited to values below the maximum allowable level specified in IEC 60601-1. In this design, the peak current does not exceed 100 µA across the intended frequency range.

To further improve performance and reduce sensitivity to resistor mismatch, the classical Howland configuration is modified using an instrumentation amplifier structure [5]. This approach improves current stability, reduces common-mode errors, and simplifies the overall circuit implementation. Two identical current source stages are used to generate a symmetric excitation suitable for four-wire bioimpedance measurements.

C. Differential Voltage Measurement

The alternating current flowing through the body produces a voltage drop between the inner electrode pair, which carries information about the tissue impedance and its temporal variations. This voltage is measured using a dedicated differential amplifier stage.

The voltage sensing amplifier must exhibit high input impedance, high common-mode rejection ratio, and sufficient gain to resolve small impedance changes caused by pulsatile blood flow. Based on these requirements, an instrumentation amplifier configuration is employed.

The input stage of the AD5933 can be modeled as an inverting amplifier with its noninverting input biased at half the supply voltage. The gain of the voltage sensing path is therefore selected to ensure that the amplified signal remains within the allowable input range of the AD5933 while preserving measurement sensitivity. The overall system gain can be expressed as:

$$g = -\frac{R_{fb}}{R_{in}} \cdot PGA\left(1 + \frac{19{,}8 \cdot 10^3}{R_G}\right) \tag{2}$$

where R_{fb} and R_{in} define the interface to the AD5933 input, PGA is the internal programmable gain of the AD5933, and R_G sets the gain of the instrumentation amplifier. This configuration allows flexible adjustment of sensitivity depending on the measured impedance range.

D. Calibration and Switching Network

Accurate impedance measurement requires compensation for gain errors, phase offsets, and frequency dependent characteristics of the analog front end. For this purpose, a digitally controlled analog switch is used to alternate between a known calibration impedance and the unknown device under test.

Calibration is performed by measuring a precision resistor at the same excitation frequency as the subsequent bioimpedance measurement. The measured real and imaginary components are stored and used to compute the magnitude and phase of the unknown impedance according to standard ratio-metric calibration equations.

For continuous measurements at a fixed frequency, calibration is performed once prior to data acquisition. This approach provides stable and repeatable impedance results while keeping system complexity low.

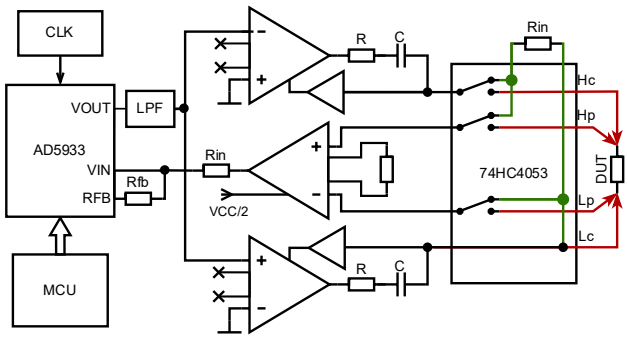


Fig. 1. Block diagram of designed analog frontend with calibration (green) and output (red) path highlighted

E. Results

The proposed bioimpedance measurement system was evaluated using precision resistors, a skin equivalent RC phantom, and in vivo measurements. The objective was to verify measurement accuracy, frequency response, and compliance with safety requirements.

Impedance magnitude was measured for precision resistors ranging from 50 Ω to 300 Ω. For resistances above 100 Ω, the relative error remained below 2% across the tested frequency

range. Higher deviations were observed at lower impedances due to reduced voltage amplitude at constant excitation current. Phase measurements showed stable behavior within the operational bandwidth, with increasing error at higher frequencies.

To assess frequency dependent performance, a simplified skin equivalent RC circuit was tested [6]. The measured impedance magnitude and phase closely matched the simulated model, confirming the system's ability to resolve both resistive and capacitive components and to track the expected decrease in impedance with increasing frequency.

A proof-of-concept in vivo measurement was performed using four electrodes spaced 20 mm apart on the forearm. A 16 kHz excitation signal was used, and impedance was sampled at 300 Hz. After low-pass filtering (50 Hz cutoff), the waveform exhibited periodic fluctuations corresponding to the cardiac cycle. As shown in Fig. 2, the bioimpedance signal follows the morphology of the reference optical PPG signal, with clearly distinguishable systolic peaks. The estimated heart rate was approximately 58 bpm, demonstrating that the system can detect bioimpedance variations associated with arterial pulse activity.

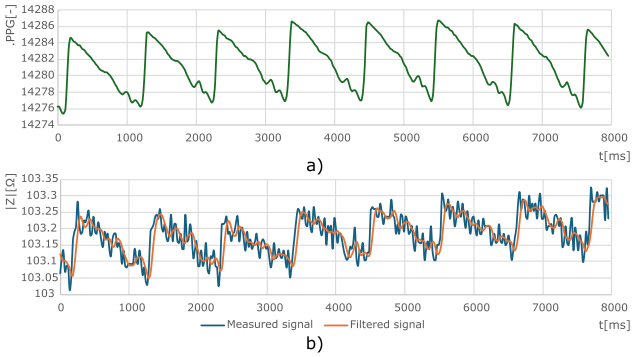


Fig. 2. Comparison of measured pulse wave with a) optical PPG and b) bioimpedance sensor.

IV. Sensing electrodes

The measurements presented in the previous sections were performed using commercial wet Ag/AgCl electrodes. While these electrodes provide low and stable interface impedance under controlled conditions, their performance may degrade over time due to gel drying and changes at the electrode-skin interface. For this reason, such electrodes are not well suited for long-term monitoring applications [7].

To address this limitation, dry electrodes have been developed for use in future long-term bioimpedance monitoring systems. The fabricated dry electrode designs are shown in Fig. 3. Screen-printed electrodes were produced on flexible substrates including PEN, Mylar, Kapton, and paper (Fig. 3. a - d), while flexible PCB-based electrodes with ENIG and immersion silver surface finishes are shown in Fig. 3. e - f. Additional surface treatments are under investigation to improve electrode-skin contact impedance and long-term stability.

In future work, the previously described developed analog front end will be used to experimentally compare the performance of these dry electrodes with standard wet Ag/AgCl electrodes. The ultimate goal is to integrate both the sensing electrodes and measurement electronics onto a single flexible substrate, enabling a fully integrated and wearable bioimpedance monitoring system.

To enable safe and repeatable testing, a human forearm phantom is currently under development. The phantom will consist of a silicone structure with an internal channel simulating a blood vessel. A stepper motor-controlled pump will circulate a blood mimicking fluid through the system, allowing controlled simulation of pulsatile blood flow. This setup will provide a stable and reproducible environment for evaluating electrode performance and system sensitivity under dynamic conditions.

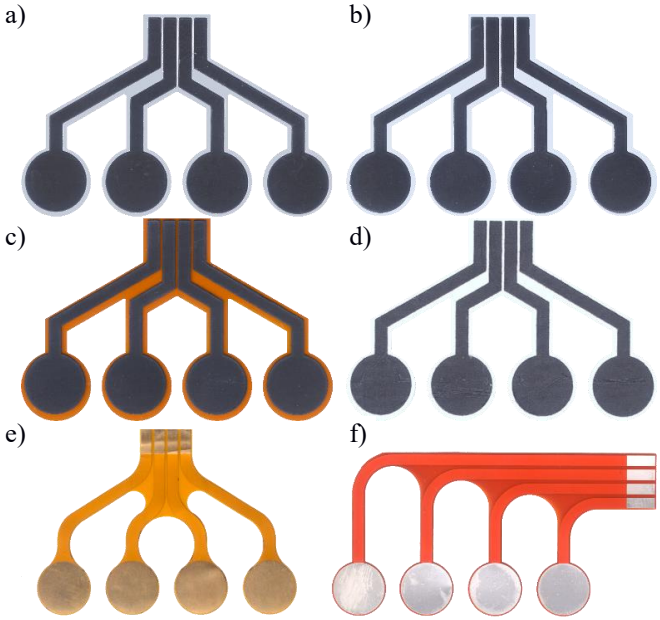


Fig. 3. Produced electrode on a) PEN, b) Mylar, c) Kapton, d) paper substrate and on flexible PCB with e) ENIG and f) immersion silver surface finish.

REFERENCES

[1] T. K. Bera, "Bioelectrical Impedance Methods for Noninvasive Health Monitoring: A Review," *J Med Eng*, vol. 2014, p. 381251, 2014, doi: 10.1155/2014/381251.

[2] S. Abasi, J. R. Aggas, G. G. Garayar-Leyva, B. K. Walther, and A. Guiseppi-Elie, "Bioelectrical Impedance Spectroscopy for Monitoring Mammalian Cells and Tissues under Different Frequency Domains: A Review," *ACS Meas. Sci. Au*, vol. 2, no. 6, pp. 495–516, Dec. 2022, doi: 10.1021/acsmeasuresciau.2c00033.

[3] *Medical electrical equipment. Part 1, General requirements for basic safety and essential performance*, Edition 3.2, Consolidated version. Geneva, Switzerland: International Electrotechnical Commission, 2020.

[4] M. Jusko and P. Lukacs, "Design of Analog Front End for Measuring Bioimpedance," in *2025 International Spring Seminar on Electronics Technology (ISSE)*, Budapest, Hungary: IEEE, May 2025, pp. 1–5. doi: 10.1109/ISSE65583.2025.11120984.

[5] I. V. Lam, "Analysis of Improved Howland Current Pump Configurations," 2020.

[6] S. I. Birlea, G. J. Corley, N. M. Birlea, P. P. Breen, F. Quondamatteo, and G. OLaighin, "Detecting electroporation by assessing the time constants in the exponential response of human skin to voltage controlled impulse electrical stimulation," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN: IEEE, Sep. 2009, pp. 1355–1358. doi: 10.1109/IEMBS.2009.5334144.

[7] N. Afif, S. Pratama, F. Haryanto, S. Khotimah, and S. Suprijadi, "Comparison of Wet and Dry EEG Electrodes Based On Brain Signals Characterization In Temporal and Anterior Frontal Areas Using Audio Stimulation," *Journal of Physics: Conference Series*, vol. 1505, p. 012069, Mar. 2020, doi: 10.1088/1742-6596/1505/1/012069.

Using visualizations to increase teaching effectiveness in the field of algorithms and data structures

[1]*Nazarii Herasymchuk (1st year),*
Supervisor: [2]*Slavomír Šimoňák*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]nazarii.herasymchuk@student.tuke.sk, [2]slavomir.simonak@tuke.sk

***Abstract*—This work focuses on the development of algorithms and data structures visualizations which represent an important approach for supporting computer science education, particularly in areas that are highly abstract and difficult for students to understand using traditional teaching methods. The aim of this work is to analyze the current state of algorithm and data structure visualization tools used in education, identify their benefits and limitations, and highlight open problems that persist despite ongoing technological advances in this field.**

***Keywords*—Algorithm visualization, data structures, interactive learning, computer science education, educational technologies**

I. Introduction

For a long time, one of the most challenging aspects of computer science education has been the teaching of algorithms and data structures. The abstract nature of the concepts involved and the high cognitive demands imposed on students—particularly in comprehending dynamic processes and intricate interactions between data elements—are the main causes of this difficulty. A thorough mental grasp of algorithmic behavior and the internal workings of data structures is frequently not possible using traditional teaching methods, which are usually dependent on textual explanations, pseudocode, and static graphics.

In response to these challenges, research in the field of algorithm and data structure visualization has significantly expanded in recent years. Numerous studies indicate that visualizations and animations can enhance comprehension, increase student motivation, and support active learning. However, despite these promising findings, several open issues remain. Visualization tools frequently encourage passive observation rather than meaningful engagement, offer limited adaptability to different levels of prior knowledge, and lack integration with immediate feedback.

These issues suggest that the mere availability of visualization tools is not sufficient [1]. The key challenge lies in designing systems that are not only visually illustrative but also interactive, adaptive, and pedagogically effective. This creates opportunities for further research focused on integrating artificial intelligence, enhancing interactivity, enabling comparative visualizations, and leveraging modern web-based technologies. The following analysis examines selected open problems in algorithm and data structure visualization and outlines potential directions for future research.

II. Literature Review and Analysis

Algorithm and data structure visualization research has evolved from simple animation tools toward more interactive and pedagogically oriented systems. Although numerous platforms demonstrate the educational value of visualization, several conceptual and technical limitations remain. The following analysis reviews key research directions, evaluates existing approaches, and identifies areas where further development is both necessary and feasible.

A. Interactive Exercises and Evaluation of Understanding

One of the most frequently cited limitations of visualization systems is the risk of passive observation. Hansen et al. [2] and Naps et al. [3] emphasize that visualization alone does not guarantee learning; active engagement is essential.

Interactive visualization environments can address this limitation by requiring students to perform algorithmic steps themselves. Three levels of interactivity can be distinguished:

Guided step execution, where students manually perform algorithm steps and receive immediate feedback.

Input manipulation and experimentation, allowing students to modify input data and observe changes in algorithm behavior.

Full implementation tasks, enabling students to write their own algorithmic solutions.

While the third level offers high pedagogical value, it significantly increases system complexity. A balanced approach combining structured interactive tasks with automated feedback appears most feasible and educationally effective.

The importance of structured interaction is further demonstrated by Perháč and Šimoňák [4], who proposed an interactive visualization system combining animation with guided exercises and evaluation mechanisms.

Research by Šimoňák (2020) [5], as well as recent empirical studies [6], [7], indicates that interactive visualization tools produce significantly better academic outcomes compared to passive animation-based tools. Nevertheless, adaptive task difficulty remains insufficiently explored. Future systems should incorporate performance-based progression mechanisms to adjust complexity according to user competence.

B. Comparative Visualization and Input Scenario Analysis

Understanding algorithm efficiency requires analyzing behavior under different input conditions (best, average, and

worst case). However, many visualization tools present only a single execution trace without enabling systematic comparison.

Comparative visualization systems such as ALGOCOMPARE [8] demonstrate the pedagogical potential of displaying multiple algorithms simultaneously. This approach enables direct observation of differences in execution flow, structural transformations, and time complexity behavior.

Similarly, dynamic graph algorithm visualization research [9] shows that visual comparison enhances structural reasoning and comprehension of algorithmic processes. Despite these advances, comparative tools remain limited in scope and often lack adaptive features or integrated assessment components.

Future research should investigate how synchronized multi-algorithm visualizations influence conceptual transfer and abstraction skills.

C. Adaptive Visualization Based on User Expertise

Learners differ significantly in prior knowledge and cognitive capacity. Static visualization systems fail to account for these differences. Adaptive educational systems aim to tailor content complexity, explanation detail, and animation speed to the user's expertise level.

Pedagogical recommendations by Šimoňák (2014) [10] and Végh (2023) [11] emphasize the importance of adjusting visualization depth according to learner experience. Beginners benefit from slower animations and detailed commentary, whereas advanced students prefer concise, technically dense representations.

Despite strong theoretical support, most current visualization libraries do not implement systematic adaptation mechanisms. This gap represents a promising research opportunity, particularly when combined with AI-based user modeling techniques.

D. Integration of Artificial Intelligence into Algorithm Visualizations

The integration of artificial intelligence (AI) into educational visualization systems represents a significant emerging research direction. Traditional algorithm visualization tools provide animated representations of algorithm execution but typically lack contextual explanations and adaptive instructional support. Recent research suggests that combining visualization with AI-driven conversational agents can enhance student engagement and conceptual understanding [12], [13].

Two main technical approaches can be identified. The first involves training a domain-specific model tailored to algorithm education. Although this approach enables full customization, it requires extensive datasets, computational resources, and expertise in machine learning, making it less feasible for an individual research project. The second, more practical solution is the integration of an existing large language model through an API interface. Such models already possess broad knowledge of algorithms and programming concepts and can generate explanations dynamically based on the current state of the visualization.

Empirical evidence presented by Bhandari (2025) [13] demonstrates that combining algorithm visualization with an AI chatbot improves perceived understanding and learning efficiency. However, systematic experimental validation comparing AI-augmented systems with traditional visualization tools remains limited. Future research should therefore focus on controlled studies evaluating measurable learning outcomes, cognitive load, and long-term knowledge retention.

The integration of AI also raises pedagogical design questions, including how to structure prompts, how to ensure explanation accuracy, and how to prevent over-reliance on automated assistance.

E. Virtual Reality and Spatial Visualization

Virtual reality (VR) introduces the possibility of spatial representations of data structures, particularly trees and graphs. Three-dimensional interaction may improve comprehension of hierarchical and relational structures.

Although existing research primarily focuses on 2D web-based systems, work on dynamic graph exploration [9], [14] and structural modeling in educational visualization [11] suggests that immersive environments may enhance spatial reasoning.

However, VR implementation introduces technical complexity, hardware dependency, and usability challenges. Rigorous evaluation is required to determine whether VR provides measurable educational advantages over advanced 2D interactive systems.

F. Extension of Existing Platforms vs. Development of a New Library

This section contrasts two realistic development paths for algorithm visualization: extending existing platforms and building a new modular visualization library. Extending established tools—such as Algomaster [15] and ALGOCOMPARE [8]—can be attractive because it leverages mature systems and may reduce development time. In practice, however, the integration of more advanced functionality (e.g., flexible modular composition, adaptive learning logic, synchronized multi-algorithm comparison, or AI-assisted contextual explanation) often requires substantial architectural changes that are difficult to implement cleanly in tools that were not designed with these requirements in mind.

A useful reference point is the JSAV library [16], which remains one of the best-known frameworks for algorithm animation and interactive exercises. Its stability and pedagogical grounding are clear strengths. At the same time, it reflects older web design assumptions, offers limited modularity, and does not provide native support for adaptive control layers or AI-based assistance [17]. Similar constraints tend to appear across many long-standing visualization systems, where new capabilities can be added only at the cost of increasing complexity and weakening maintainability.

For these reasons, this thesis primarily follows the second path: the design of a modular prototype, understood as a supporting visualization core/library rather than a full-scale replacement of existing solutions. Developing a new core makes it possible to adopt modern architectural patterns and integrate more naturally with contemporary frontend frameworks, while also providing a clean foundation for comparative visualization modes, adaptation mechanisms, and AI-assisted explanation modules. Depending on implementation goals, the rendering layer can be built on technologies such as D3.js or WebGL-based frameworks to support both efficient 2D interactions and higher-performance visualizations.

At the same time, the aim is not to compete with mature platforms through feature parity. A complete replacement would be resource-intensive and unrealistic within dissertation constraints. Instead, the prototype serves as an experimental platform that demonstrates how extensibility, adaptability, and AI integration can be treated as first-class design goals and then evaluated in a controlled manner. This direction is justified not as a purely technological exercise, but as a response to identifiable pedagogical limitations in current systems and the difficulty of integrating these capabilities through incremental extensions alone.

G. Feasibility Constraints and Architectural Implications

While several directions appear conceptually promising, not all of them are equally feasible within realistic individual research constraints. For example, fully immersive virtual reality environments for algorithm visualization introduce significant hardware dependence, usability challenges, and evaluation complexity, making rigorous validation difficult without dedicated experimental infrastructure. Similarly, training a custom domain-specific AI model would require large annotated datasets and substantial computational resources, which is typically beyond the scope of a dissertation prototype.

For this reason, the most practical and research-valuable direction lies in improving existing visualization paradigms through a modular architecture that enables incremental extension. A key limitation of many established tools is that they were designed around older architectural assumptions, which can restrict modularity, integration with modern frontend frameworks, and systematic implementation of adaptive logic layers. This often makes advanced features—such as learner modeling, structured progression, synchronized multi-algorithm comparison, or AI-assisted explanation—hard to integrate without major re-engineering.

Consequently, rather than aiming to replace mature platforms, a more defensible strategy is to design a prototype modular core that demonstrates how structured interactivity, adaptation mechanisms, and AI-based contextual explanation can coexist within a single extensible framework. Such a prototype can serve as an experimental platform for controlled evaluation while remaining realistic in implementation effort. In addition, broader survey work emphasizes that visualization's educational effectiveness depends strongly on pedagogical integration and engagement design, not on visualization availability alone [1].

III. Problem Identification and Future Directions

The conducted analysis of current research and existing visualization platforms confirms that substantial progress has been achieved in the field of algorithm and data structure visualization. Numerous studies demonstrate that visual representations improve student motivation, conceptual understanding, and engagement when compared to purely textual or lecture-based instruction [10], [5], [6], [7]. Interactive systems such as JSAV [16] have successfully incorporated guided exercises and automated assessment mechanisms, while platforms like Algomaster [15] and ALGOCOMPARE [8] extend visualization capabilities toward improved usability and comparative algorithm analysis.

These developments indicate that the foundational problem—whether visualization contributes positively to algorithm education—has been convincingly addressed. The effectiveness of interactive visualization, when pedagogically integrated, is supported by empirical evidence. However, the analysis also reveals that the current generation of tools solves only part of the broader instructional challenge.

One persistent issue is the risk of passive engagement. Although many systems include animations and limited interaction controls, a significant portion of visualization tools still operate primarily as demonstration environments. Learners frequently observe algorithm execution without being required to actively construct reasoning steps. Research consistently emphasizes that observation alone does not guarantee conceptual understanding. While some platforms integrate interactive tasks, systematic frameworks for structured progression and deeper cognitive engagement remain underdeveloped.

Adaptivity and personalization represent a second unsolved issue. Regardless of prior knowledge or skill level, existing systems usually apply a standard educational paradigm to all users. While expert learners benefit from condensed representations and less explanatory overhead, beginners might need slower execution, thorough explanations, and scaffolded supervision. Dynamic adaptation based on user performance is rarely used, despite compelling pedagogical grounds for differentiated training. Instead of using automated learner modeling, the majority of visualization solutions rely on static settings or human parameter change.

A third limitation relates to intelligent instructional support. Although recent research explores the integration of artificial intelligence into educational systems [12], [13], the majority of algorithm visualization platforms lack contextualized explanation mechanisms. Students who encounter misunderstanding must rely on external resources or instructor intervention. The potential of AI-assisted explanation—particularly through large language models capable of generating natural-language clarifications—remains largely untapped in mainstream visualization frameworks. Moreover, rigorous comparative studies measuring objective learning gains from AI-augmented visualization systems are still limited.

In addition to pedagogical limitations, technical and architectural constraints restrict further evolution of existing tools. Libraries such as JSAV [16] provide stable and pedagogically validated environments but are built upon older architectural paradigms that limit modularity and extensibility. Native support for adaptive logic layers, AI integration, or scalable comparative multi-algorithm visualization is not inherently embedded within these systems. As a result, extending their functionality often requires substantial modification rather than straightforward modular expansion.

Not all identified issues are equally feasible to address within the scope of doctoral research. For instance, the development of fully immersive virtual reality environments for algorithm visualization would require extensive hardware resources, user testing, and long-term validation studies. Similarly, training a custom domain-specific artificial intelligence model would demand large annotated datasets and considerable computational infrastructure. While these directions are conceptually promising, they exceed realistic individual research constraints.

Considering both pedagogical impact and implementation feasibility, the most promising direction for further research lies in the design of an integrated modular visualization framework that addresses three key dimensions simultaneously.

First, the system should implement structured interactivity combined with automated evaluation. Instead of merely displaying algorithm animations, the environment should require students to actively perform algorithmic steps, make decisions, and predict outcomes. Immediate feedback mechanisms must be incorporated to reinforce correct reasoning and correct misconceptions at early stages.

Second, the framework should include an adaptive visualization layer. User interaction data and task performance metrics can be used to dynamically adjust explanation depth, animation speed, and task complexity. Such adaptation does not require full-scale artificial intelligence but can initially rely on rule-based progression models. Over time, this layer may evolve toward more sophisticated learner modeling approaches.

Third, the system should integrate an AI-based explanatory assistant, implemented through an existing large language model API. Rather than replacing instructional structure, the AI component would function as a contextual support tool capable of generating clarifications linked to the current visualization state. This approach avoids the complexity of training a custom model while enabling dynamic explanation capabilities. Careful prompt engineering and system constraints would be necessary to ensure explanation consistency and reliability.

Importantly, the proposed direction does not aim to completely replace existing visualization platforms. A more feasible strategy involves the development of a prototype modular core that demonstrates improved architectural flexibility, adaptability, and AI integration. Such a prototype would serve as an experimental research platform rather than a fully commercial system.

From an educational perspective, future research should focus on how visualization can support active learning rather than passive observation. The main goal should not be only to improve the technical design of visualization tools, but to create learning environments in which students actively solve algorithmic tasks, receive immediate feedback, and build a deeper understanding of algorithm behavior and data structure operations. In this sense, the future direction lies in strengthening the pedagogical role of visualization as an instructional aid that supports reasoning, practice, and gradual knowledge construction.

A promising direction is therefore the development of a learning-oriented visualization environment that combines three educational principles: active student engagement, adaptation to the learner's current level of knowledge, and timely explanatory support. Such a system should guide students through algorithmic processes step by step, adjust the form and complexity of visualization according to performance, and provide additional explanations when difficulties arise. This would make visualization not only a demonstration tool, but also a means of individualized learning support.

Overall, the future of algorithm and data structure visualization should be understood primarily as an educational challenge rather than only a technical one. The most valuable contribution of further research will be to design and verify solutions that help students learn more effectively, understand concepts more deeply, and engage more actively in the learning process. In this way, visualization can evolve from a supportive display technology into a more comprehensive pedagogical instrument within computer science education.

References

[1] E. Fouh, M. Akbar, and C. A. Shaffer, "The role of visualization in computer science education," Virginia Tech, Technical Report, 2014, survey paper on algorithm visualization in computer science education. [Online]. Available: https://people.cs.vt.edu/shaffer/cs6604/Papers/AVpedagogypost.pdf

[2] S. Hansen, N. H. Narayanan, and M. Hegarty, "Designing educationally effective algorithm visualizations," *Journal of Visual Languages and Computing*, vol. 13, no. 3, pp. 291–317, 2002. [Online]. Available: https://doi.org/10.1006/jvlc.2002.0236

[3] T. Naps, G. Rößling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, M. Mcnally, and S. Rodger, "Exploring the role of visualization and engagement in computer science education," *ACM SIGCSE Bulletin*, vol. 35, pp. 131–152, 08 2002. [Online]. Available: https://doi.org/10.1145/782941.782998

[4] P. Perháč and S. Šimoňák, "Interactive system for algorithm and data structure visualization," *Computer Science Journal of Moldova*, vol. 30, pp. 28–48, 02 2022. [Online]. Available: https://doi.org/10.56415/csjm.v30.02

[5] S. Šimoňák, "Increasing the engagement level in algorithms and data structures course by driving algorithm visualizations," *Informatica*, vol. 44, 09 2020. [Online]. Available: https://doi.org/10.31449/inf.v44i3.2864

[6] G. Kogan, H. Chassidim, and I. Rabaev, "The efficacy of animation and visualization in teaching data structures: a case study," *Educational Technology Research and Development*, pp. 1–24, 06 2024. [Online]. Available: https://doi.org/10.1007/s11423-024-10382-w

[7] Y. Zhang, "Application of algorithm visualization techniques in teaching computer data structure course," *Applied Mathematics and Nonlinear Sciences*, vol. 9, 04 2024. [Online]. Available: https://doi.org/10.2478/amns-2024-0788

[8] F. Vateha and S. Šimoňák, "Comparative visualization of algorithms and data structures," *Computing and Informatics*, vol. 44, pp. 336–365, 04 2025. [Online]. Available: https://doi.org/10.31577/cai_2025_2_336

[9] M. Burch, H. Wetering, G. Wallner, F. Rooks, and O. Morra, "Exploring the dynamics of graph algorithms," *Journal of Visualization*, vol. 26, 10 2022. [Online]. Available: https://doi.org/10.1007/s12650-022-00885-0

[10] S. Šimoňák, "Using algorithm visualizations in computer science education," *Open Computer Science*, vol. 4, 10 2014. [Online]. Available: https://doi.org/10.2478/s13537-014-0215-4

[11] L. Végh, "Models of data structures in educational visualizations for supporting teaching and learning algorithms and computer programming," *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 7, pp. 147–154, 06 2023. [Online]. Available: https://doi.org/10.59287/ijanser.916

[12] G. Madejski, R. Madejski, and W. J. Kulesza, "AI-driven tool supporting algorithmic thinking in programming education," in *2025 7th Experiment@ International Conference (exp.at'25)*, 2025, pp. 373–377. [Online]. Available: https://doi.org/10.1109/exp.at2565440.2025.11347115

[13] M. Bhandari, "Algorithm visualization on data structures along with AI chatbot," *International Journal for Research in Applied Science and Engineering Technology*, vol. 13, pp. 605–612, 06 2025. [Online]. Available: https://doi.org/10.22214/ijraset.2025.72158

[14] H. Ed Dien and Y. Asnar, "OPT+Graph: Detection of graph data structure on program visualization tool to support learning," in *2018 International Conference on Data and Software Engineering (ICODSE)*, 11 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICODSE.2018.8705794

[15] M. Benej and S. Šimoňák, "Algomaster platform extension for improved usability," *Journal of Electrical and Electronics Engineering*, vol. 10, pp. 27–30, 05 2017.

[16] V. Karavirta and C. Shaffer, "Creating engaging online learning material with the JSAV JavaScript algorithm visualization library," *IEEE Transactions on Learning Technologies*, vol. 9, pp. 1–1, 10 2015. [Online]. Available: https://doi.org/10.1109/TLT.2015.2490673

[17] G. Rößling, "A first set of design patterns for algorithm animation," *Electronic Notes in Theoretical Computer Science*, vol. 224, pp. 67–76, 2009, proceedings of the Fifth Program Visualization Workshop (PVW 2008). [Online]. Available: https://doi.org/10.1016/j.entcs.2008.12.050

A Multimodal Deep Learning Framework for Detecting Hoaxes and Fake Images in the Digital Environment

[1]*Miroslav MURIN(3rd year),*
Supervisor: [2]*Miroslav MICHALKO*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]miroslav.murin@tuke.sk, [2]miroslav.michalko@tuke.sk

***Abstract*—This paper addresses the growing threat of hoaxes, fake news and AI-manipulated media, with a special focus on their impact on children and young people. It summarises recent evidence on the rapid rise of AI-generated child sexual abuse material and non-consensual synthetic imagery, as well as the broader role of hoaxes in political manipulation and online fraud. In response, we propose a multimodal detection system that combines EfficientNetV2 for real-versus-fake image classification, a face-analysis module for deepfake detection, and transformer-based language models such as RoBERTa for identifying phishing and hoax-like text, extended to video via frame sampling and synthetic-speech analysis. By aggregating the outputs of these components into a unified risk score, the architecture aims to offer a robust, practically deployable tool for identifying harmful content and supporting safer digital environments, particularly in educational and youth-oriented contexts.**

***Keywords*—AI-generated images, Cyberbullying, Deepfakes, Fake news, Hoaxes, Sexual abuse, Multimodal machine learning, Phishing detection**

I. Introduction

In the modern digital world, information spreads faster than ever before. Along with many positive effects, this brings new problems, especially the rise of false information and manipulated images. Today, artificial intelligence tools can easily create fake photos and videos that look very real. Because of this, it is becoming harder to tell what is true and what has been digitally created. These false materials often spread quickly on social media and can seriously damage the reputation of individuals or groups.

Young people are one of the most at-risk groups in this environment. According to the authors in [1] and [2], one of the most pressing issues today is the safety of our children in cyberspace. The world of the internet hides many serious threats that can negatively affect the emotional, social, and mental health of children. In schools, fake or edited images are increasingly used to hurt or embarrass others. Sometimes such pictures have a sexual tone or are designed to make someone a target of public ridicule. For students, these situations can cause strong emotional distress, loss of confidence, and lasting psychological harm. The fast and anonymous nature of the internet makes the problem even worse, as false images can reach a wide audience within minutes. The authors in [3] emphasize the importance of relevant online educational programs aimed at engaging teenagers and adults in critical thinking that could help in detecting fake news. The authors in [4] argue that comprehensive interventions are needed to limit the creation, sharing, and consumption of non-consensual sexual content, including synthetic content.

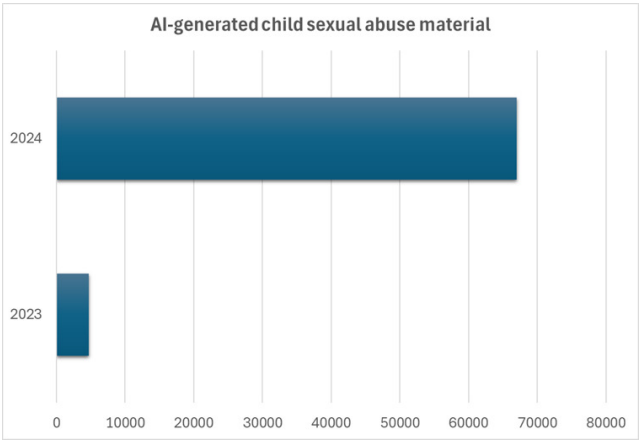


Fig. 1. AI-generated child sexual abuse material is rising

According to the authors in [5], the number of cases of child sexual abuse generated by artificial intelligence is rising: between 2023 and 2024, the number of reports increased by 1,325%, with growing concerns about deepfake images in which children's faces are placed on sexual material. This increase was recorded in reports to the National Center for Missing and Exploited Children, which rose from 4,700 reports recorded in 2023 to more than 67,000 in 2024, as we can see in the figure 1.

According to the author in [6], in the digital age, fake news poses a serious threat, especially during elections, as it undermines unity and causes conflict. With millions of fake news stories circulating, the risk is further heightened by the public's susceptibility to fake news, often fueled by emotive language and a lack of verifiable sources. According to the authors in [7], a new era of information warfare has dawned. Various actors, including those supported by the state, are using information on social networks as a weapon to conduct disinformation campaigns aimed at manipulating public opinion on specific issues. These disinformation campaigns can have disastrous consequences for the public: they can change its opinions and actions. At the same time, hoaxes are frequently connected to various forms of fraud, for

example through fake charity campaigns, phishing schemes or fabricated investment opportunities, which exploit people's fear, curiosity or lack of information. In this way, hoaxes contribute to an environment of confusion and distrust, where it becomes increasingly difficult for citizens to distinguish credible information from manipulation and deception.

II. Proposed solution for detecting hoaxes and fake images

A possible way to address these challenges is to design and train a set of machine learning models that work together as a coordinated system. Instead of relying on a single classifier, the proposed solution uses several specialized models, each focused on a different aspect of the problem—for example, detecting visual manipulation in images, identifying misleading textual patterns in accompanying captions or posts, and analysing audio. By combining their outputs, these models can provide a more robust and reliable assessment of whether a given image or piece of content is likely to be a hoax or a fake. This ensemble approach aims to reduce false positives and false negatives, and to adapt more easily to new types of attacks and emerging manipulation techniques. Figure 2 shows how text and image (left) or video (right) analysis is performed.

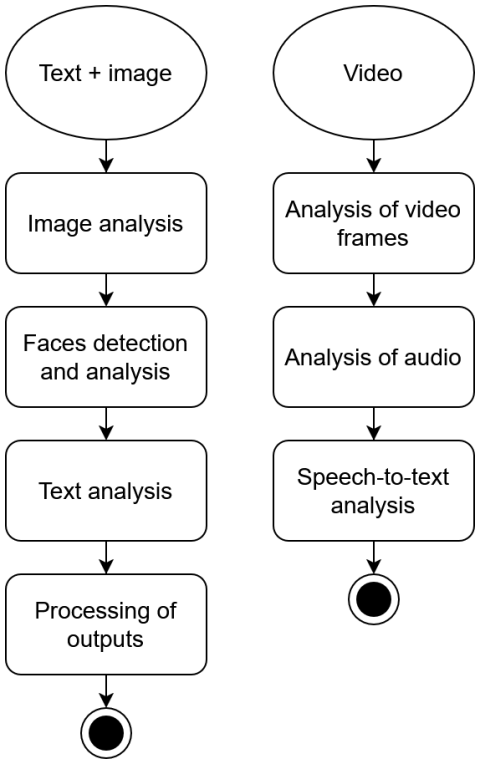


Fig. 2. Diagram analyzing potential hoaxes. On the left is an analysis of images and text (e.g., a social media post), and on the right is a video analysis.

A. Analysis of text and images

A robust approach to detecting dangerous images and hoaxes is to combine multiple models, allowing the weaknesses of individual components to be compensated by the strengths of others. In this system, a general image classifier is used to determine whether an image is real or AI-generated. It is trained on large-scale datasets containing both authentic photographs and synthetic images produced by generative models such as Stable Diffusion, DALL·E-like systems, Midjourney, and other GAN- or diffusion-based approaches. However, since many manipulation techniques modify only specific parts of an image, especially faces, relying solely on a global image analysis is often insufficient.

To address this, a second model focuses specifically on facial analysis. The system first detects and extracts all faces present in the image and then evaluates each cropped face using a specialised classifier to determine whether it is genuine or synthetically altered. This enables the detection of deepfake-style manipulations where only identities are modified while the rest of the image remains visually consistent. Such a targeted approach improves sensitivity to subtle artefacts that may not be captured by a general image classifier.

The final component of the system performs text analysis. This stage can be divided into two models: one for detecting hoax-like or misleading content and another for identifying phishing or fraud-related messages. When text is available, it is analysed together with the image to capture both semantic and contextual signals of deception; otherwise, this step is skipped. The outputs of all models are then aggregated into a unified risk score, allowing the system to classify content as benign or potentially harmful, including fully generated images, partially manipulated images, and genuine images combined with misleading text.

B. Analysis of video

Building on this approach, video analysis is performed in several stages. First, a subset of frames is extracted at predefined intervals to reduce computational cost. Each selected frame is analysed using the same image-based model as for still images, estimating the probability of manipulation. The final visual score is obtained by computing the median of these predictions, which helps reduce the impact of occasional misclassifications.

However, videos may be visually authentic while containing manipulated audio. Therefore, an additional module evaluates whether the audio track is natural or synthetically generated based on characteristic acoustic artefacts. The audio is also transcribed using automatic speech recognition and analysed by text-based models for phishing and hoax detection. By combining visual, audio and textual analysis, the system provides a comprehensive assessment of whether a video represents potentially harmful or deceptive content.

III. Training models

For classifying real versus fake images, we can employ the EfficientNetV2 architecture, which consistently ranks among the top performers in image classification benchmarks while remaining extremely fast and compact. According to the authors in [8] EfficientNetV2 was designed to optimise both accuracy and computational efficiency, achieving up to 6.8× smaller model size and 5–11× faster training compared to previous state-of-the-art models on ImageNet. When pretrained on large-scale datasets such as ImageNet-21k, EfficientNetV2 models reach higher top-1 accuracy than recent Vision Transformer variants (e.g. ViT-L/16), despite using fewer parameters and requiring less computation. These properties make EfficientNetV2 a suitable backbone for our system, where high

classification accuracy must be combined with fast inference in real-time or near-real-time applications.

For training the model that detects whether texts are phishing or hoax messages, we can employ RoBERTa (Robustly Optimized BERT Approach), a transformer-based language model that improves on the original BERT architecture through a refined pre-training procedure. According to the authors in [9] RoBERTa keeps the same encoder structure as BERT but removes the next-sentence prediction objective, uses much larger mini-batches, trains on significantly more data (over 160 GB of text), and applies dynamic masking so that different tokens are masked in different epochs. The base variant, RoBERTa-base, uses 12 transformer encoder layers with 768-dimensional hidden states and about 110 million parameters, which makes it well suited as a backbone for fine-tuning on classification tasks including phishing detection, spam filtering and automatic identification of hoax-like or misleading text content.

IV. Experiments

In the experimental phase, we trained a text-based model for hoax detection. At the current stage, however, its performance is still insufficient for reliable deployment, as it achieves an accuracy of only 68.57%. This result indicates that the text component still struggles to distinguish hoax-like messages from benign content with the desired level of robustness. In future work, we plan to improve this part of the system by extending the training dataset, refining the preprocessing pipeline, and experimenting with alternative transformer-based architectures and fine-tuning strategies.

On the other hand, the face-manipulation detection model achieved very promising results. In a test conducted on 70 images, the model reached an accuracy of 92.86%, a precision of 0.9167, a recall of 0.9429, an F1-score of 0.9296, and an AUC-ROC of 0.9616. These values indicate that the classifier is able to distinguish real and fake faces with high reliability and very good separability between the two classes. The confusion matrix showed only five incorrect predictions in total: three real images were misclassified as fake, and two fake images were misclassified as real. The class-wise results were also well balanced, with precision/recall values of 0.9412/0.9143 for the real class and 0.9167/0.9429 for the fake class. This suggests that the model does not strongly favor one class over the other and is effective in detecting manipulated faces while maintaining a low number of false predictions. Figure 3 shows the Confusion matrix and Confidence Histogram from the 70-face detection test.

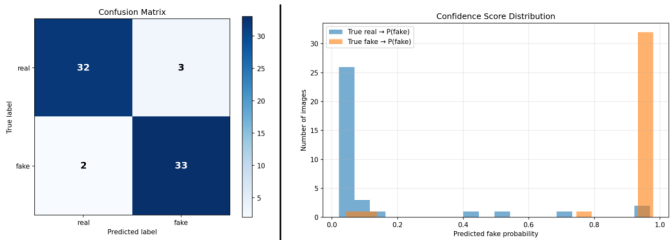


Fig. 3. Confusion matrix and Confidence Histogram from face detection model testing

V. Conclusion

In conclusion, the rapid spread of hoaxes, fake news and AI-manipulated media poses a growing threat to individuals and to society as a whole, with children and young people among the most vulnerable. Synthetic sexual content, deepfake images and misleading narratives can seriously damage victims' emotional well-being, distort public debate and undermine trust in democratic institutions and information sources.

This work proposes a modular detection system that combines several specialised machine learning models to analyse images, video, audio and text in a unified way. By integrating an efficient image classifier such as EfficientNetV2 with face-focused deepfake analysis and transformer-based language models like RoBERTa for phishing and hoax detection, the approach aims to reliably identify a wide range of harmful content while remaining suitable for practical deployment.

Experimental results show that while the text-based hoax detection model currently achieves limited performance (68.57% accuracy) and requires further improvement, the face manipulation detection component performs very well, reaching an accuracy of 92.86% along with high precision, recall and AUC-ROC values, demonstrating strong potential for practical use.

Such tools, when complemented by education and media literacy, can form an important part of a broader strategy to reduce the impact of hoaxes and fake images in the digital environment.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 17R05-04-V01-00004, Competence and Training Center for Cyber and Information Security TUKE.

References

[1] S. A. Laczi and V. Póser, "Impact of deepfake technology on children: Risks and consequences," in *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, 2024, pp. 215–220.

[2] A. J. Sánchez-Medina, I. Galván-Sánchez, and M. Fernández-Monroy, "Applying artificial intelligence to explore sexual cyberbullying behaviour," *Heliyon*, vol. 6, no. 1, p. e03218, 2020. [Online]. Available: https://doi.org/10.1016/j.heliyon.2020.e03218

[3] A. Maftei, A.-C. Holman, and I.-A. Merlici, "Using fake news as means of cyber-bullying: The link with compulsive internet use and online moral disengagement," *Computers in Human Behavior*, vol. 127, p. 107032, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563221003551

[4] R. Umbach, N. Henry, R. Shelby, G. Stevens, and K. Gonzalez-Pons, "Ai-generated image-based sexual abuse: Perpetration and consumption across three regions," *Computers in Human Behavior*, p. 108935, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563226000324

[5] D. Fry, "Millions of children face sexual violence as ai deepfakes drive surge in new cases—latest global data," *The Conversation*, 10 2025. [Online]. Available: https://theconversation.com/millions-of-children-face-sexual-violence-as-ai-deepfakes-drive-surge-in-new-cases-latest-global-data-266171

[6] B. Juarto, "Classification of hoax news using machine learning and neural networks with bert embeddings," in *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2023, pp. 311–316.

[7] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, "The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans," *J. Data and Information Quality*, vol. 11, no. 3, May 2019. [Online]. Available: https://doi.org/10.1145/3309699

[8] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," *CoRR*, vol. abs/2104.00298, 2021. [Online]. Available: https://arxiv.org/abs/2104.00298

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized BERT pretraining approach," *CoRR*, vol. abs/1907.11692, 2019. [Online]. Available: http://arxiv.org/abs/1907.11692

Deep learning methods for natural language processing

[1]*Samuel FODOR (1st year),*
Supervisor: [2]*Peter BEDNÁR*

[1,2]Institute of Artificial Intelligence, FEEI, Technical University of Košice, Slovak Republic

[1]samuel.fodor@tuke.sk, [2]peter.bednar@tuke.sk

***Abstract*—Deep learning has transformed natural language processing by enabling end-to-end models that learn language representations from large corpora and achieve state-of-the-art performance on a wide range of NLP tasks. Recent work has proposed knowledge fusion methods that aim to combine the capabilities of several trained models into a single system, usually by directly merging parameters, composing modular adapters, or distilling knowledge into smaller students. This paper presents a structured overview of deep learning methods for NLP with a focus on weight-space merging and related model-reuse techniques. First, it outlines the foundations of pretrained language models and the rise of large pretrained models. Second, it presents an overview of fusion approaches as they are applied to large language models. Finally, it summarizes common empirical patterns and open questions, as well as possible directions for future research.**

***Keywords*—Knowledge fusion, Large language models, Model merging, Natural language processing.**

I. Introduction

Deep learning has transformed natural language processing by enabling end-to-end models that learn language representations from large corpora and achieve state-of-the-art performance on a wide range of NLP tasks. Large language models based on the transformer architecture now serve as general-purpose text generators and reasoning engines, but their size and training cost make it difficult to adapt for many real-world scenarios.

An increasingly prominent challenge is how to obtain models that are simultaneously specialized, efficient, and easy to deploy, without repeatedly training or fine-tuning models or networks from scratch. Recent work has proposed knowledge fusion methods that aim to combine the capabilities of several trained models into a single model, usually by directly merging parameters, composing modular adapters, or distilling knowledge into smaller students.

This paper reviews deep learning methods for NLP with a focus on knowledge fusion in weight space and related model-reuse techniques. First, it outlines the foundations of deep learning for NLP and the rise of large pretrained models. Second, it presents a structured overview of knowledge fusion methods, covering direct weight merging, structure-aware merging, modular fusion, and distillation-based approaches as they are applied to large language models. Finally, it summarizes common empirical patterns and open questions, as well as possible directions for future research.

II. Deep learning for NLP and model reuse

A. Pretrained large language models and fine-tuning

Modern NLP pipelines are dominated by large pretrained models that learn generic language representations from massive corpora of text, which are then adapted to downstream tasks via supervised fine-tuning [1]. Early work on BERT-style masked language models and GPT-style autoregressive models has evolved into multi-billion-parameter LLMs that are capable of instruction following, few-shot learning, and multi-task generalization across a wide range of tasks and domains [2], [3].

Fine-tuning remains the standard way to specialize an LLM to a new task or domain, but it raises multiple potential issues. These issues range from catastrophic forgetting when models are fine-tuned sequentially, to the need to store many separate checkpoints for different tasks, and high computational costs of full-parameter updates [4], [5]. These limitations have incentivized research into more efficient adaptation methods and mechanisms, with the aim of reusing existing models rather than training new ones for every setting [6], [7].

B. Parameter-efficient adaptation and modularity

Parameter-efficient fine-tuning techniques such as adapters and low-rank adaptation attach small task-specific modules to a frozen base model, enabling specialization with only a small fraction of the parameters being trained [6]. This makes it viable to maintain many task or domain-specific heads while sharing a common base model, and several works demonstrate that trained adapters can reach performance comparable to full-parameter fine-tuning [2], [1].

PEFT methods also enable a modular approach to knowledge fusion, where multiple adapters trained on different data sets or tasks are combined, weighted, or dynamically selected to unify their strenghts without modifying the base model weights [6]. This modular view complements parameter-space merging of full models and offers an cost-efficient method for deep learning methods that must operate under memory and computational constraints while still making use of various sources of expertise in NLP [1].

C. Ensembles and distillation

Although beyond the scope of single-model adaptation, classical approaches to combining models in NLP include ensembling, where several models' predictions are aggregated

at inference time, and knowledge distillation, where a smaller student model is trained to match the behavior of one or more teacher models [2]. Ensembles can improve accuracy and robustness but increase inference time and cost, while distillation offers compression, but typically requires additional training data and optimization [2], [8].

Recent work extends distillation to LLMs by aligning not only output distributions, but also intermediate modules and attention maps, aiming to transfer internal structures to smaller students and improve generalization on NLP tasks [2]. At the same time, research has begun to explore training-free or training-light parameter-space merging techniques as an alternative way to combine LLMs into single models that integrate multiple tasks or domains [4].

III. Knowledge fusion methods for LLMs

A. Direct weight averaging and task-vector methods

The simplest category of weight-space fusion methods uses linear interpolation between models that share the same architecture and initialization. Basic approaches include global weight averaging and linearly combining task vectors, to form new models that aim to inherit skills from several parent models. Task vectors are defined as the differences between fine-tuned and base checkpoints [4]. Empirical studies show that appropriately scaled linear combinations can transfer capabilities between tasks or domains, for example by adding a domain-specific task vector to a generic instruction-tuned model [4], [6], [9].

More advanced techniques introduce non-linear interpolation such as spherical linear interpolation (SLERP) between checkpoints, or sparsity and regularization-based schemes like TIES and DARE that select or rescale only subsets of the parameter deltas to reduce interference [4], [9], [10]. These methods have been used to merge medically oriented LLMs with general-purpose chat models, resulting in single models that perform competitively on both medical benchmarks and general reasoning tasks without additional full-parameter fine-tuning [7].

Task arithmetic and related operations also offer a conceptual link between model merging and linear representations of "skills", suggesting that some aspects of NLP behavior can be treated as approximate directions in parameter space [4], [1], [3]. However, the resulting performance is sensitive to both scaling and to the overlapping or conflicting nature of the target tasks, and naive averaging across many tasks often massively degrades results. This motivates more structured or data-driven approaches [11], [12].

B. Structure-aware and search based merging

Because not all layers contribute to different tasks equally, recent work has proposed layer or group-wise merging strategies that assign distinct mixing coefficients across depth [13], [14]. One such strategy is KAFA-Merge, which views merging two fine-tuned LLMs as an optimization problem over layer-group blending ratios and uses Bayesian optimization followed by local linear search to find effective configurations [11]. By searching in a reduced space of layer groups and evaluating on small data subsets, KAFA-Merge achieves improved performance compared to uniform global merging while remaining computationally viable.

Other structure-aware approaches rely on neuron or head alignment, permutation matching, or optimal transport to align internal units before averaging. This is especially useful when models have undergone different fine-tuning trajectories [4], [15]. These methods attempt to ensure that semantically similar neurons are merged together, reducing interference that can occur when misaligned units are averaged. They are specifically relevant for transformer-based LLMs in NLP, where attention heads and MLP blocks may specialize in different linguistic or reasoning patterns, but alignment adds complexity and often requires additional computation [4], [15], [1].

A complementary line of work treats merging as a sequential decision process and uses reinforcement learning to discover layer-wise "recipes". Reinforced Model Merging, for instance, defines actions that choose among copying a layer from a parent model, applying a merging operator, or skipping or revisiting layers, and trains an agent to maximize downstream performance using approximate rewards on small subsets of data [16]. These approaches can include many existing operators and search over architectures of fused models, showing gains on NLP benchmarks while reducing the number of full evaluations required [1], [16].

C. Modular and adapter-based fusion

Instead of merging full sets of weights, several works investigate fusion at the level of modular components. In these scenarios, a single frozen base LLM hosts multiple sets of trainable low-rank modules, each trained on a different dataset or task, and the challenge becomes how to combine these adapters into a unified behavior [6], [7], [9]. A typical setup trains separate LoRA heads for categories of data (like chat, question answering, domain-specific instructions), and then learns scalar fusion weights that mix the outputs or updates from each adapter [6].

Fusion of LoRA adapters has been shown to integrate multiple instruction datasets into one model without re-running full fine-tuning on the union, mitigating issues like dataset conflict that can arise when all data are mixed from the start [6]. Compared to merging full checkpoints, adapter fusion can be easier to control, since the base model remains fixed and only the small adapters are combined. However, it still requires careful tuning of fusion weights and may not capture deep interactions between tasks [6], [9].

Modular fusion ideas also appear in methods that assemble or route between experts at inference time, for example mixture-of-experts architectures or systems that dynamically select among specialized LLMs based on the input [1]. While these approaches often require some additional training or routing infrastructure, they share the goal of weight-space merging in reusing existing models to construct richer NLP systems without fully retraining models from scratch [4], [1].

D. Distillation and multi-teacher fusion

Although knowledge distillation does not merge weights from multiple parents directly, it is closely related to knowledge fusion. Recent work on distillation for LLMs proposes to match not only logits but also intermediate representations and attention patterns between teacher and student, improving the transfer of semantic and contextual information across tasks

[2]. Experimental results indicate that such multi-level distillation can produce students that approach teacher performance while using significantly fewer parameters.

Multi-teacher (or knowledge amalgamation) methods extend this idea by training a student model on several teachers at once. Teachers may be specialized in different tasks or domains, and the student learns to reproduce their behavior on corresponding data, with the aim of absorbing the knowledge of multiple tasks or domains into a single model. As mentioned, unlike weight merging, this requires a fresh training run, but it avoids the issues with merging heterogeneous architectures, since here knowledge fusion happens through losses rather than direct arithmetic on weights. [17], [2].

IV. Patterns in knowledge fusion for NLP

Across the reviewed works, we noted several recurring patterns regarding the effectiveness and limitations of knowledge fusion methods applied to NLP models.

A. Benefits compared to sequentially fine-tuned baselines

Many fusion methods reliably outperform the original pre-trained base models on the union of tasks for which fine-tuned parents are available. Simple averaging or task-vector approaches can already yield improvements, especially when parents are well aligned and trained from the same initialization. More sophisticated methods (like SLERP) or sparsity-aware operators often close much of the gap to individual task-specialized models [4], [9]. This effect is especially visible when merging domain-adapted models into general-purpose LLMs, as in biomedical or low-resource language settings [7], [11].

Fusion can offer advantages over sequential fine-tuning by mitigating catastrophic forgetting. Instead of overwriting weights while fine-tuning on tasks one by one, merging combines checkpoints trained on different data, preserving performance on earlier tasks more effectively in multiple reported experiments [4], [10]. Some methods make use of this property to build multi-domain or multi-task LLMs from separate fine-tuned parents without requiring access to all training data at once, which is beneficial when data is private or distributed [16].

Modular fusion of adapters and LoRA components allows the integration of several instruction datasets or specialized behaviors into a single model with relatively little additional parameter cost [6]. This is particularly relevant for NLP applications where many small datasets capture different styles, domains, or safety constraints, and where storing a full checkpoint per dataset would be prohibitive [6], [8], [18].

B. Remaining gaps to single-task and multi-task baselines

Despite these benefits, fused models typically underperform strong single-task models that have been fine-tuned directly on the target data, especially on challenging or highly specialized NLP tasks [9]. FusionBench and related studies observe that multi-task training from scratch or from a shared base model still sets a high bar that current training-free merging methods do not consistently reach [9], [16].

In addition, performance is sensitive to the nature of the tasks being merged. When tasks are closely related or share linguistic phenomena, merging tends to work well, whereas when tasks or domains are more heterogeneous, naive merging can lead to negative interactions that reduce performance on all tasks involved compared to their fine-tuned parents [11], [16]. This has led to the use of search-based and reinforcement learning approaches that adapt layer-wise recipes to the specifics of each task combination, but these methods introduce additional complexity and hyperparameters that must be tuned [16].

Even in adapter-based fusion, the choice of fusion weights and the compatibility of datasets heavily influence performance. For example, combining many heterogeneous instruction datasets via a single joint fine-tuning can hurt performance relative to training and fusing task-specific adapters. This suggests that fusion is not merely a matter of adding more data, but requires careful control of how different signals interact [6], [18]. From the perspective of deep learning methods for NLP, this shows that knowledge fusion in its current form is not a drop-in replacement for established training approaches and must be applied with an understanding of task relationships and model properties.

C. Compression and deployment

A final pattern relates to the role of knowledge fusion in model compression and deployment. Feature-alignment based distillation shows that students can inherit much of the performance of teachers while using fewer parameters and being easier to deploy. This is especially relevant for NLP models serving on edge devices or in resource-constrained environments [2]. At the same time, merging methods such as those implemented in MergeKit and demonstrated in domain-specific systems like BioMistral showcase how weight-space fusion can be combined with quantization and PEFT to obtain specialized yet efficient LLMs [4], [7].

V. Future research

A. Methodological and evaluation challenges

A prominent open question is how to design consistent benchmarks and evaluation protocols for knowledge fusion methods in NLP. FusionBench is a step toward standardizing comparisons across tasks and models, but its scope spans both vision and language, and there is room for more focused benchmarks that target NLP-specific phenomena such as syntactic understanding, discourse, and long-context reasoning [9]. Toolkits like MergeKit have accelerated experimentation, yet many published studies rely on custom datasets or task combinations, making it difficult to draw general conclusions across works [4]. This suggests high value in constructing or adopting a unified benchmark that evaluates different fusion strategies across a controlled set of NLP tasks and model families.

Another methodological challenge regards hyperparameter selection and search strategies for merging. Approaches such as KAFA-Merge and RMM demonstrate that layer-wise coefficients and operator choices can have a large impact on performance, but they also introduce additional search spaces that must be explored efficiently [11], [16]. Open questions include how to automate or regularize these searches, how to design constraints that reflect linguistic structure, and how to balance global and local decisions along the depth of a transformer. For deep learning methods in NLP, this points to the need for optimization frameworks that make fusion methods more robust and less dependent on manual tuning.

B. Understanding representation and transfer in parameter space

Outside of practical concerns, there are conceptual questions about what weight-space operations reveal about how LLMs represent linguistic knowledge. Task vectors and successful merges suggest that some capabilities can be approximated and manipulated linearly, while failures and negative interactions show that many aspects of model behavior remain insufficiently explored [4], [11]. There is ongoing work on analyzing which layers or heads are most responsible for particular tasks, but relatively little is known about how these structures change under merging or adapter fusion, and how this relates to generalization in NLP tasks [10], [7].

VI. Conclusion

This work has reviewed deep learning methods for NLP with a focus on knowledge fusion techniques that combine the capabilities of multiple models into a single system. Starting from the foundations of pretrained language models and parameter-efficient adaptation, it has shown how weight-space merging, modular fusion, and knowledge distillation offer complementary ways to reuse and combine existing models for new tasks and domains. The reviewed works demonstrate that fusion methods can significantly improve over pretrained baselines and offer practical advantages compared to maintaining many separately fine-tuned models, especially when combined with quantization and PEFT for deployment.

At the same time, current fusion techniques generally do not fully match the best task-specific or multi-task fine-tuning baselines and require careful design of merging strategies, hyperparameters, and modular compositions. This leaves room for research on better benchmarks, optimization methods, and deeper understanding of how parameter-space operations affect representations and transfer in NLP models. In the context of deep learning methods for natural language processing, these open questions define promising directions for exploring how knowledge fusion can become a practically useful tool for building adaptable cost-effective NLP solutions, as well as exploring methods of quantifying the impact of fusion on base models.

References

[1] E. Yang, L. Shen, G. Guo, X. Wang, X. Cao, J. Zhang, and D. Tao, "Model merging in llms, mllms, and beyond: Methods, theories, applications, and opportunities," *ACM Comput. Surv.*, vol. 58, no. 8, Feb. 2026. [Online]. Available: https://doi.org/10.1145/3787849

[2] S. Wang, C. Wang, J. Gao, Z. Qi, H. Zheng, and X. Liao, "Feature alignment-based knowledge distillation for efficient compression of large language models," 2024. [Online]. Available: https://arxiv.org/abs/2412.19449

[3] Y. Zhou, L. Song, B. Wang, and W. Chen, "MetaGPT: Merging large language models using model exclusive task arithmetic," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 1711–1724. [Online]. Available: https://aclanthology.org/2024.emnlp-main.102/

[4] C. Goddard, S. Siriwardhana, M. Ehghaghi, L. Meyers, V. Karpukhin, B. Benedict, M. McQuade, and J. Solawetz, "Arcee's MergeKit: A toolkit for merging large language models," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, F. Dernoncourt, D. Preoţiuc-Pietro, and A. Shimorina, Eds. Miami, Florida, US: Association for Computational Linguistics, Nov. 2024, pp. 477–485. [Online]. Available: https://aclanthology.org/2024.emnlp-industry.36/

[5] M. Tao, C. Zhang, Q. Huang, T. Ma, S. Huang, D. Zhao, and Y. Feng, "Unlocking the potential of model merging for low-resource languages," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 8705–8720. [Online]. Available: https://aclanthology.org/2024.findings-emnlp.508/

[6] D. Gao, Y. Ma, S. Liu, M. Song, L. Jin, W. Jiang, X. Wang, W. Ning, S. Yu, Q. Xuan, X. Cai, and L. Yang, "Fashiongpt: Llm instruction fine-tuning with multiple lora-adapter fusion," *Know.-Based Syst.*, vol. 299, no. C, Sep. 2024. [Online]. Available: https://doi.org/10.1016/j.knosys.2024.112043

[7] Y. Labrak, A. Bazoge, E. Morin, P.-A. Gourraud, M. Rouvier, and R. Dufour, "BioMistral: A collection of open-source pretrained large language models for medical domains," in *Findings of the Association for Computational Linguistics: ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 5848–5864. [Online]. Available: https://aclanthology.org/2024.findings-acl.348/

[8] H. A. A. K. Hammoud, U. Michieli, F. Pizzati, P. Torr, A. Bibi, B. Ghanem, and M. Ozay, "Model merging and safety alignment: One bad model spoils the bunch," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 13 033–13 046. [Online]. Available: https://aclanthology.org/2024.findings-emnlp.762/

[9] A. Tang, L. Shen, Y. Luo, E. Yang, H. Hu, L. Zhang, B. Du, and D. Tao, "Fusionbench: A unified library and comprehensive benchmark for deep model fusion," 2025. [Online]. Available: https://arxiv.org/abs/2406.03280

[10] A. Alexandrov, V. Raychev, M. N. Müller, C. Zhang, M. Vechev, and K. Toutanova, "Mitigating catastrophic forgetting in language transfer via model merging," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 17 167–17 186. [Online]. Available: https://aclanthology.org/2024.findings-emnlp.1000/

[11] M. K. Yuce and M. F. Amasyali, "Kafa-merge: Model merging method with layer-based bayes+linear search," *IEEE Access*, vol. 13, pp. 108 743–108 755, 2025.

[12] G. Perin, X. Chen, S. Liu, B. Kailkhura, Z. Wang, and B. Gallagher, "RankMean: Module-level importance score for merging fine-tuned LLM models," in *Findings of the Association for Computational Linguistics: ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 1776–1782. [Online]. Available: https://aclanthology.org/2024.findings-acl.104/

[13] G. Du, J. Lee, J. Li, R. Jiang, Y. Guo, S. Yu, H. Liu, S. K. Goh, H.-K. Tang, D. He, and M. Zhang, "Parameter competition balancing for model merging," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, ser. NIPS '24. Red Hook, NY, USA: Curran Associates Inc., 2024.

[14] Y. Ju, Z. Ni, X. Xing, Z. Zeng, H. Zhao, S. Fan, and Z. Zhang, "Mitigating training imbalance in LLM fine-tuning via selective parameter merging," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 15 952–15 959. [Online]. Available: https://aclanthology.org/2024.emnlp-main.892/

[15] N. Verma and M. Elbayad, "Merging text transformer models from different initializations," 2024. [Online]. Available: https://arxiv.org/abs/2403.00986

[16] J. Han, J. Ye, S. Liu, H. Zhang, J. Song, Z. Feng, and M. Song, "Reinforced model merging," 2025. [Online]. Available: https://arxiv.org/abs/2503.21272

[17] Z. Lu, C. Fan, W. Wei, X. Qu, D. Chen, and Y. Cheng, "Twin-merging: dynamic integration of modular expertise in model merging," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, ser. NIPS '24. Red Hook, NY, USA: Curran Associates Inc., 2024.

[18] T. Boros, R. Chivereanu, S. Dumitrescu, and O. Purcaru, "Fine-tuning and retrieval augmented generation for question answering using affordable large language models," in *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*, M. Romanyshyn, N. Romanyshyn, A. Hlybovets, and O. Ignatenko, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 75–82. [Online]. Available: https://aclanthology.org/2024.unlp-1.10/

Quantifying Student Contributions in Collaborative Programming Projects Through Advanced Git-Based Metrics

[1]*Tomáš BUČEK (3rd year),*
Supervisor: [2]*Ján GENČI*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]tomas.bucek@tuke.sk, [2]jan.genci@tuke.sk

***Abstract*—Assessing individual contributions in team-based programming courses remains a persistent challenge in higher education. Traditional evaluation approaches often rely on subjective peer assessment or coarse indicators such as total commits and lines of code, which fail to capture the depth, structure, and collaborative nature of student work. This paper proposes a data-driven methodology for analyzing Git repository data in order to quantify individual student contributions in collaborative software development assignments.**

We introduce a set of experimental composite metrics derived from commit-level metadata, file-level changes, and temporal activity patterns. In addition to basic indicators (e.g., commits and executable lines of code), we propose novel measures such as the Code Impact Score, Sequential Complexity Score, and Collaborative Integration Score. These metrics are normalized and combined into an overall contribution score adaptable to specific course objectives.

Future work focuses on integrating artificial intelligence techniques for automated interpretation of repository data and validating the approach through real-world deployment in team-based semester projects. The ultimate goal is to establish a practical and interpretable evaluation model that enhances both assessment quality and learning outcomes in collaborative programming education.

***Keywords*—Git, Student Contribution Metrics, Educational Data Mining**

I. Introduction

Collaborative software development has become a standard component of higher education in computer science and related disciplines. In university programming courses, students are increasingly required to work in teams, often using distributed version control systems such as Git to manage their code. While teamwork promotes practical skills, communication, and real-world development experience, it also introduces challenges in objectively assessing individual contributions within group projects. Instructors frequently face difficulties in determining how much work each student has actually performed, particularly in larger teams or long-term semester projects. Ensuring fairness, transparency, and accountability in grading therefore becomes a non-trivial problem.

Version control repositories provide a detailed and structured record of development activity, including commits, code modifications, file changes, and commit messages. These artifacts offer a potential basis for quantitative analysis of student contributions. However, simplistic metrics such as the number of commits or total lines of code modified may fail to reflect true effort or learning outcomes. Minor formatting changes can inflate activity counts, while complex debugging or architectural design decisions may not correspond to large code differences. Consequently, there is a need for a systematic and balanced evaluation framework that combines multiple repository-derived indicators into a composite contribution score. In this paper, we propose a weighted metric model tailored to educational settings, aiming to support instructors in assessing individual effort within collaborative programming assignments in a more objective and data-driven manner.

II. Related Work

The use of version control data to assess individual contributions in software development projects has been explored in various contexts. Early work on repository mining established foundational metrics such as commit frequency and distribution, highlighting how commit patterns can reflect developer activity and engagement within open source projects [1]. Repository mining techniques have subsequently been applied to quantify developer productivity and contribution, emphasizing the challenges of interpreting simple metrics without contextual information [2].

Within the educational domain, several studies have investigated how Git and associated metrics can support the assessment of student participation in team-based programming projects. Hamer et al. [3] employed Git contribution metrics to analyze inequality and participation patterns among students in undergraduate software engineering courses, demonstrating the potential of repository metrics as instructional feedback tools. Similarly, Quesada López et al. [4] examined the impact of individual commits on software quality during student projects by performing commit-impact analysis, suggesting that mining repository data can reveal insights into both quality and performance learning outcomes.

Research [5] in educational practice has also evaluated how presenting Git activity metrics to students impacts teamwork and course outcomes. For instance, integrating GitLab-derived activity metrics into coursework consultation sessions was shown to help students identify uneven workload distribution and improve engagement. Beyond the classroom, empirical frameworks [6] have been proposed for refining contribution rates to mitigate anomalies and better capture sustained development effort, which could be extended to academic settings.

Furthermore, broader studies on developer metrics in open source and collaborative environments inform the selection and interpretation of contribution indicators. Şeker et al. [7] discussed new developer metrics in the context of open source software challenges, highlighting the need for multi-dimensional assessment. In parallel, analytical tools that extract fine-grained collaboration networks from commit data illustrate how detailed repository analysis can support understanding of team dynamics [8]. Collectively, this body of work motivates the development of composite contribution metrics tailored to educational settings, where both fairness in assessment and interpretability are paramount.

III. Git Data Extraction and Representation

Version control systems such as Git provide a detailed record of software development activities, which can be leveraged to analyze individual contributions in team-based programming courses. Each Git repository contains a history of commits, branches, merges, and tags, along with metadata such as author, timestamp, commit message, and file changes. In the context of student projects, these data allow instructors to objectively quantify who contributed what, when, and where, enabling fairer assessment of individual effort within group assignments.

A. Storage and Access

Git stores each commit as an object identified by a SHA-1 hash. Commits reference parent commits and contain snapshots of the project tree, representing all files at the time of the commit. Metadata such as the author, date, and commit message are stored with each object. Programmatically, these data can be accessed via commands like `git log` or through libraries such as GitPython, which allow iteration over all commits and extraction of relevant statistics including lines added, lines deleted, and modified files.

For example, a single commit authored by the instructor might be represented in JSON as shown in Listing 1.

Listing 1: Example JSON representation of a Git commit

```
{
  "commit_hash": "a1b2c3d4e5f6g7h8i9j0",
  "author": "tomas.bucek@tuke.sk",
  "date": "2026-02-15T14:32:00Z",
  "message": "Fixed bug",
  "files_changed": [
    {
      "file_path": "input_parser.py",
      "lines_added": 15,
      "lines_deleted": 3
    },
    {
      "file_path": "parser.py",
      "lines_added": 5,
      "lines_deleted": 0
    }
  ]
}
```

This structured format allows for straightforward aggregation and analysis of student activity across the entire project. Key indicators such as executable lines of code, number of files touched, and bug-fix commits can be computed from this representation to quantify individual effort.

B. Extraction Workflow and Practical Considerations

A typical workflow for analyzing student repositories involves cloning the repository, iterating over all branches and commits, and collecting metrics per student. Derived metrics include total commits, lines added and deleted, modified files, and detected bug-fix commits (identified by keywords in commit messages). These metrics can then be aggregated to produce a dataset for the entire course, enabling calculation of productivity, ownership, and overall contribution scores.

Practical considerations include normalizing author identities (to account for multiple emails or usernames used by a student), filtering out merge commits or automated formatting commits, and noting that not all effort (such as design discussions or offline coding) is captured. Despite these limitations, Git data provides a highly detailed and objective view of each student's activity, supporting instructors in assessing individual contributions in collaborative programming assignments.

IV. Methodology: Quantifying Student Contributions from Git Data

In this work, we propose a set of composite metrics designed to capture not only the volume of student activity in programming courses, but also the depth, complexity, and collaborative impact of their contributions. These metrics are derived solely from Git repository data, including commit metadata, file-level changes, and commit messages. In addition to standard indicators such as total commits and executable lines of code, we introduce novel experimental metrics tailored to educational settings.

A. Base Metrics

For each student, the following base metrics are extracted from the repository:

- C: total number of commits
- LA: total lines added
- LD: total lines deleted
- $NC = LA - LD$: net code contribution
- FT: number of unique files modified
- MC: number of unique modules (top-level directories) touched
- BF: number of bug-fix commits (detected by keywords such as `fix`, `bug`)

B. Experimental Contribution Metrics

Beyond traditional metrics, we propose three new composite metrics, which were designed purely for experimental purpose, and following evaluation.

1) Code Impact Score (CIS): While executable lines of code measure quantity, they do not capture the relative importance of the files modified. To address this, we define the Code Impact Score as:

$$CIS_i = \frac{1}{|F_i|} \sum_{f \in F_i} \log(1 + LOC_f) \cdot W_f$$

where F_i is the set of files modified by student i, LOC_f is the executable lines changed in file f, and W_f is a weighting factor representing the total number of students who modified the file (files modified by fewer students are weighted higher to reflect ownership and initiative). This encourages recognition of substantial contributions to critical or less-touched files.

2) Sequential Complexity Score (SCS): To reward contributions that involve continuous logical development rather than scattered trivial commits, we define Sequential Complexity Score:

$$SCS_i = \frac{\sum_{k=1}^{N_i} \Delta LOC_k \cdot e^{-\Delta t_k/\tau}}{N_i}$$

where N_i is the total number of commits by student i, ΔLOC_k is the executable lines changed in commit k, Δt_k is the time elapsed since the previous commit, and τ is a decay factor controlling sensitivity to spacing between commits. Large, temporally coherent contributions yield higher SCS values.

3) Collaborative Integration Score (CIntS): To reflect contributions that meaningfully integrate with team work (e.g., resolving conflicts or touching modules others depend on), we define:

$$CIntS_i = \sum_{m \in M_i} \frac{1}{1 + \log(1 + D_m)}$$

where M_i is the set of modules touched by student i, and D_m is the number of other students who have modified the same module. This metric assigns higher scores to students who contribute to modules that are critical yet less frequently modified by others, promoting both initiative and collaborative awareness.

C. Overall Contribution Score

Finally, all metrics are normalized to percentiles $p \in [0, 1]$ across all students, and combined into an overall score:

$$Overall_i = w_1 \cdot p(CIS_i) + w_2 \cdot p(SCS_i) + w_3 \cdot p(CIntS_i) + w_4 \cdot p(BF_i)$$

Weights w_1, w_2, w_3, w_4 are configurable to balance emphasis on code impact, sequential development, collaborative integration, and bug fixing. This allows instructors to tailor the evaluation framework to the learning objectives of a specific programming course.

D. Example Calculation

Using the JSON commit example in Listing 1, we can illustrate a simple calculation for a single student:

- Commit `a1b2c3d4e5f6g7h8i9j0` modifies two files: `src/input_parser.py` $(+15, -3)$ and `tests/test_input_parser.py` $(+5, 0)$
- Net LOC: $NC = (15 - 3) + (5 - 0) = 17$
- CIS: since each file is modified by only this student, $W_f = 1$, then $CIS = \log(1 + 12) \cdot 1 + \log(1 + 5) \cdot 1 \approx 2.56 + 1.79 = 4.35$
- SCS: assuming time difference $\Delta t = 2$ hours and $\tau = 24$ hours, $SCS = 17 \cdot e^{-2/24} \approx 15.8$
- $CIntS$: touching two modules with $D_m = 0$ and 1 respectively: $CIntS = 1/(1+\log(1)) + 1/(1+\log(2)) = 1 + 0.59 = 1.59$
- $BF = 1$ (keyword `fix` detected)

After computing percentiles across all students and applying weights (e.g., $w_1 = 0.3$, $w_2 = 0.25$, $w_3 = 0.25$, $w_4 = 0.2$), the `Overall` contribution score can be calculated, providing a nuanced, multi-dimensional evaluation of student effort and impact.

V. Conclusion

In this work, we have presented a methodology for quantifying individual contributions in team-based programming assignments using Git repository data. By introducing experimental metrics such as the Code Impact Score, Sequential Complexity Score, and Collaborative Integration Score, instructors can obtain a more nuanced and multi-dimensional understanding of student activity, beyond traditional commit counts or lines of code. These metrics enable fairer assessment of effort, encourage meaningful collaboration, and provide insights into both the quantity and quality of contributions in educational projects.

Future work aims to integrate artificial intelligence techniques to automate the evaluation process. Machine learning models could analyze commit patterns, message content, and file-level changes to predict not only contribution scores, but also engagement, understanding, and learning outcomes. Such automated systems would reduce the manual burden on instructors, provide timely feedback to students, and support adaptive interventions in courses where teamwork and collaborative coding are central components.

Finally, we plan to validate the proposed metrics and AI-based evaluation framework in practice during the teaching of programming courses. By applying this approach to real semester-long team projects, we aim to assess the reliability, interpretability, and educational impact of the methodology. Successful deployment could establish a new standard for data-driven assessment of student teamwork in computer science curricula, fostering both fairness and enhanced learning outcomes.

References

[1] C. Kolassa, D. Riehle, and M. A. Salim, "The empirical commit frequency distribution of open source projects," *arXiv*, 2014. [Online]. Available: https://arxiv.org/abs/1408.4978

[2] J. Lima, C. Treude, F. F. Filho, and U. Kulesza, "Assessing developer contribution with repository mining-based metrics," in *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 2015, pp. 536–540.

[3] S. A. Hamer, C. U. Quesada López, A. Martínez Porras, and M. Jenkins Coronas, "Using git metrics to measure students' and teams' code contributions in software development projects," *CLEI Electronic Journal*, vol. 24, no. 2, 2021.

[4] C. U. Quesada López, A. Martínez Porras, and M. Jenkins Coronas, "Measuring students' source code quality in software development projects through commit-impact analysis," in *Advances in Intelligent Systems and Computing*, 2021.

[5] N. Mann *et al.*, "Integrating gitlab metrics into coursework consultation sessions in a software engineering course," *Journal of Systems and Software*, vol. 167, p. 110613, 2020.

[6] V. Bishop and S. Simske, "Model contribution rate theory: An empirical examination," *arXiv*, 2024. [Online]. Available: https://arxiv.org/abs/2412.05978

[7] A. Şeker, B. Diri, and H. Arslan, "New developer metrics for open source software development challenges: An empirical study of project recommendation systems," *Applied Sciences*, vol. 11, no. 3, p. 920, 2021.

[8] C. Gote, I. Scholtes, and F. Schweitzer, "Analysing time-stamped co-editing networks in software development teams using git2net," *arXiv*, 2019. [Online]. Available: https://arxiv.org/abs/1911.09484

Decoupling Control of Five-Port MMAB Converter Based on the Newton-Raphson Method

[1]*Tomáš BASARIK (3rd year),*
Supervisor: [2]*Milan LACKO*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]tomas.basarik@tuke.sk, [2]milan.lacko@tuke.sk

***Abstract*—Modular Multi-active bridge (MMAB) converters are essential for integrating multiple power sources and loads within a single power stage. However, the inherent cross-coupling between ports often degrades dynamic performance and complicates independent control. This paper proposes a decoupling control strategy based on the Newton-Raphson method to achieve precise port isolation. The algorithm was implemented on a dual-core MCU, where the decoupling calculations were optimized to execute within 36 µs. To ensure stable operation under varying loads, a dedicated voltage control loop was designed for each port. The proposed method was experimentally validated on a five-port (5×360 W) MAB converter. Results demonstrate that the proposed strategy successfully eliminates port interference, maintaining system stability and high dynamic responsiveness across the entire operating range.**

***Index Terms*—modular multi-active bridge converter, Newton-Raphson method, decoupling control, dual-core mcu, DC-DC power conversion, microgrids**

I. Introduction

The global transition toward carbon neutrality and the reduction of greenhouse gas emissions has significantly accelerated the adoption of renewable energy systems and electric vehicles. These systems require the integration of energy storage solutions to improve performance and reliability. Consequently, power electronics are responsible for managing the bidirectional power flow between multiple sources [1]–[3].

Traditionally, these interconnections have been achieved using individual two-port DC-DC converters. This conventional architecture involves multiple power conversion stages, which increase the total mass and reduce efficiency due to cumulative conversion losses. Multi-port converters have emerged as an alternative, offering a single-stage power conversion between multiple ports. The Multi-Active Bridge (MAB) converter has received attention thanks to its high power density, galvanic isolation and bidirectional power flow capability [1], [4]–[6].

Despite their benefits, MAB converters face inherent cross-coupling. Since all ports share a transformer network, changing the operating conditions of one port also affect the power flow in the others. This interdependency transforms the control structure into a complex multi-input, multi-output system, making independent port regulation difficult and often resulting in poor dynamic response [1], [5], [7], [8].

To address this, researchers have explored two main categories of decoupling strategies: software-based and hardware-based [1]. Software-based methods involve advanced algorithms such as ideal feedforward decoupling based on gain matrix inversion [3], Model Predictive Control [4] or Active Disturbance Rejection Control [9] to observe and mitigate cross-coupling in real-time. Hardware-based approaches aim to simplify control by modifying the converter topology, such as utilizing series-resonant circuits to minimize port impedance or designing transformers with asymmetrical leakage inductances to isolate a main port [5].

This paper presents a software decoupling method for Modular Multi-Active Bridge (MMAB) converter, see Fig. 1, using the Newton-Raphson algorithm. By solving a system of nonlinear equations, the method determines phase shifts from target current values. Unlike power-based models that reduce modularity for a reduced-order Jacobian, this work employs a full-rank current-based model. This approach provides full system modularity by prioritising all ports equally. The algorithm, implemented on a dual-core MCU, achieves an execution time of 36 µs and is validated through experimental results.

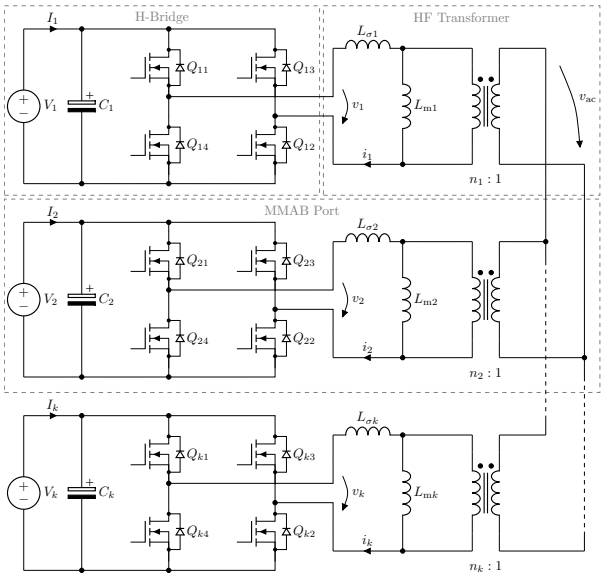


Fig. 1. Modular Multi-Active Bridge converter topology

II. Current-Based Model of MMAB Converter

Fig. 2 shows the lossless equivalent circuit of the MMAB converter. The model consists of leakage L_σ and magnetizing L_m inductances, with square voltage sources representing the H-bridges. The instantaneous AC bus voltage can be determined using Millman's theorem:

$$v_\mathrm{ac} = \frac{\sum\limits_{i=1}^{k} \frac{v_i n_i}{L_{\sigma i}}}{\sum\limits_{i=1}^{k} \left(\frac{n_i^2}{L_{\sigma i}} + \frac{n_i^2}{L_{\mathrm{m}i}} \right)} = \frac{\sum\limits_{i=1}^{k} \frac{v_i n_i}{L_{\sigma i}}}{\frac{1}{L_\mathrm{eq}}}, \tag{1}$$

where $i \in \{1, \ldots, k\}$ is the port index, v_i is the voltage, n_i is the turns ratio, $L_{\sigma i}$ is the leakage inductance of the i-th port, and L_eq is the equivalent inductance.

Using the phase-shift modulation technique, the i-th port current is defined as the sum of current contributions from the remaining ports:

$$I_\mathrm{i} = \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{V_j n_i n_j \left(\phi_i - \phi_j\right) \left(1 - 2\left|\phi_i - \phi_j\right|\right)}{f_\mathrm{s} \frac{L_{\sigma i} L_{\sigma j}}{L_\mathrm{eq}}}, \tag{2}$$

where V_j is the j-th port voltage, ϕ_i and ϕ_j are the phase shifts of the i-th and j-th ports, and f_s is the switching frequency. The individual terms in (2) can be defined as follows:

$$I_\mathrm{i} = \sum_{\substack{j=1 \\ j \neq i}}^{k} V_j K_{ij} \phi_{ij} \left(1 - 2\left|\phi_{ij}\right|\right), \tag{3}$$

where

$$K_{ij} = \frac{n_i n_j}{f_\mathrm{s} \frac{L_{\sigma i} L_{\sigma j}}{L_\mathrm{eq}}}, \qquad \phi_{ij} = \phi_i - \phi_j. \tag{4}$$

Control of the system requires identifying the specific phase shifts necessary to achieve the target port currents. However, as shown in (3), the current equations are inherently nonlinear. This means that a system of nonlinear algebraic equations must be solved, which is a computationally complex task.

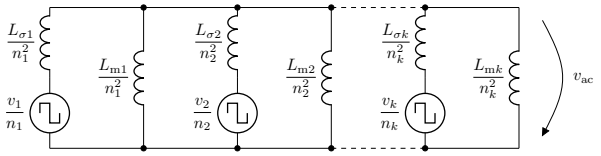


Fig. 2. Lossless equivalent circuit of MMAB Converter

III. Newton-Raphson Based Decoupling Strategy

The main objective is to determine the optimal phase-shift vector $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_\mathrm{k}]^T$ such that the output currents I_i reach their respective references $I_{i\mathrm{req}}$. The problem is formulated as a nonlinear system of equations:

$$f(\boldsymbol{\phi}) = \begin{bmatrix} I_1(\phi_1, \ldots, \phi_k) - I_{1\mathrm{req}} \\ I_2(\phi_1, \ldots, \phi_k) - I_{2\mathrm{req}} \\ \vdots \\ I_k(\phi_1, \ldots, \phi_k) - I_{k\mathrm{req}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

where $f(\boldsymbol{\phi})$ represents the current residuals (the error between actual and required currents).

The system is solved iteratively via the Newton-Raphson method, where a correction vector $\Delta\boldsymbol{\phi}$ updates the phase-shift estimates in each iteration:

$$\boldsymbol{\phi}_{z+1} = \boldsymbol{\phi}_z + \Delta\boldsymbol{\phi}_z, \tag{6}$$

where z is the iteration index. The correction vector $\Delta\boldsymbol{\phi}$ is computed using the Jacobian matrix $\mathbf{J}$ and the current residual vector $\Delta\mathbf{I}$:

$$\Delta\boldsymbol{\phi} = \mathbf{J}^{-1}(-\Delta\mathbf{I}), \tag{7}$$

expressed in full matrix form as:

$$\begin{bmatrix} \Delta\phi_1 \\ \Delta\phi_2 \\ \vdots \\ \Delta\phi_k \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial \phi_1} & \frac{\partial f_1}{\partial \phi_2} & \cdots & \frac{\partial f_1}{\partial \phi_k} \\ \frac{\partial f_2}{\partial \phi_1} & \frac{\partial f_2}{\partial \phi_2} & \cdots & \frac{\partial f_2}{\partial \phi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial \phi_1} & \frac{\partial f_k}{\partial \phi_2} & \cdots & \frac{\partial f_k}{\partial \phi_k} \end{bmatrix}^{-1} \begin{bmatrix} -\Delta I_1 \\ -\Delta I_2 \\ \vdots \\ -\Delta I_k \end{bmatrix}, \tag{8}$$

where ΔI_i represents the current residual at the i-th port, and is determined as:

$$\Delta I_i = I_i - I_{i\mathrm{req}}. \tag{9}$$

The partial derivatives that form the Jacobian matrix $\mathbf{J}$ are defined as follows:

$$\frac{\partial f_i}{\partial \phi_j} = \begin{cases} V_j K_{ij} \left(4\left|\phi_{ij}\right| - 1\right) & i \neq j, \\ \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} V_j K_{ij} \left(1 - 4\left|\phi_{ij}\right|\right) & i = j. \end{cases} \tag{10}$$

Due to Kirchhoff's Current Law, the Jacobian matrix is inherently singular, reflecting the linear dependence of port currents at the common AC bus. To address this, the Moore-Penrose pseudoinverse is employed via the algorithm in [10], providing a robust least-squares or minimum-norm solution.

While iterations typically continue until all updates $\Delta\phi_i$ fall below a threshold ε, selecting the previous phase shifts as initial conditions often enables single-iteration convergence. At minimum, this informed initial guess places the system in close proximity to the solution, drastically accelerating subsequent convergence.

The proposed decoupling algorithm, utilizing the Moore-Penrose pseudoinverse, was implemented on dual-core digital signal controler TMS320F28379D. The combined execution time for the decoupling and the control scheme described in the following section is 36 µs.

IV. Control System Design

Following system decoupling into an identity transfer matrix, the output voltage regulator is designed via polynomial-based pole placement. The control structure is illustrated in Fig. 3. For each port, the transfer function G_s is defined by the parallel combination of the output capacitance C and load resistance R:

$$G_{\mathrm{s}(s)} = \frac{R}{RCs + 1}. \tag{11}$$

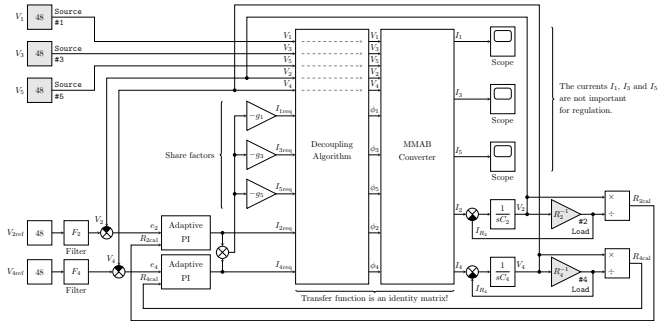


Fig. 3. Block diagram of the control system

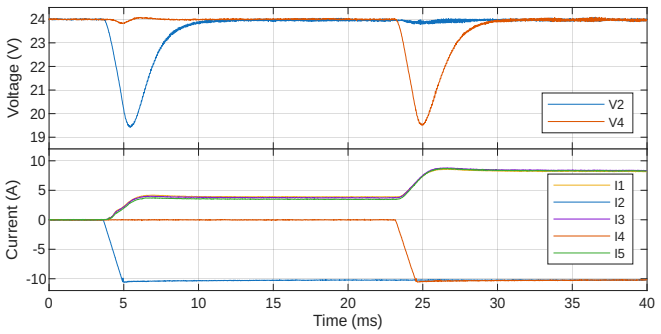


Fig. 4. Curves recorded during port 2 and 4 loading

Based on the the transfer function G_{s}, the open-loop transfer function G_{o} is formed using a proportional-integral controller. The resulting closed-loop system G_{c} is then matched to a desired second-order reference model G_{r} through coefficient comparison, where ω_0 and ζ represent the natural frequency and damping ratio, respectively. This yields the following expressions:

$$G_{\mathrm{o}(s)} = \frac{r_0 s + r_1}{s} G_{\mathrm{s}(s)} = \frac{r_0 R s + r_1 R}{RCs^2 + s}, \tag{12}$$

$$G_{\mathrm{c}(s)} = \frac{G_{\mathrm{o}(s)}}{1 + G_{\mathrm{o}(s)}} = \frac{\dfrac{r_0}{C} s + \dfrac{r_1}{C}}{s^2 + \dfrac{r_0 R + 1}{RC} s + \dfrac{r_1}{C}}, \tag{13}$$

$$G_{\mathrm{r}(s)} = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \tag{14}$$

The resulting proportional and integral gains, r_0 and r_1, are determined as follows:

$$r_0 = \frac{2\zeta\omega_0 RC - 1}{R}, \qquad r_1 = \omega_0^2 C. \tag{15}$$

Substituting the control gains (15) into (13) yields a closed-loop transfer function with an additional zero. To suppress this derivative effect, the following input filter is implemented:

$$G_{\mathrm{f}}(s) = \frac{\dfrac{r_1}{C}}{\dfrac{r_0}{C} s + \dfrac{r_1}{C}} = \frac{1}{\dfrac{r_0}{r_1} s + 1}. \tag{16}$$

Furthermore, since the load resistance R directly affects the proportional gain, the controller is updated in real-time, resulting in an adaptive PI control scheme.

V. Experimental Result

The proposed control strategy was validated on a five-port (5×360 W) MMAB converter. The experimental setup comprised three source ports #1, #3, #5 with current-sharing factors set to $g_1 = g_3 = g_5 = 1/3$ and two load ports #2 and #4. To evaluate decoupling performance, step load changes of 10 A were applied sequentially to ports #2 and #4 while monitoring the cross-channel voltage transients.

The voltage regulators were tuned for a critical damping ratio $\zeta = 1$ and a natural frequency $\omega_0 \approx 1160$ rad/s, targeting a 5 ms regulation time. The experimental waveforms, illustrating the precise decoupling and dynamic response, are presented in Fig. 4.

VI. Conclusion

A 5-port MMAB converter was decoupled using the Newton-Raphson method and Moore-Penrose pseudoinversion to resolve singularities. The DSP implementation achieved a 36 µs execution time. Experimental results showed peak deviations of 104 mV (0.43%) on V_4 and 94 mV (0.39%) on V_2 during cross-loading, with a regulation time under 5 ms. This confirms the feasibility of real-time independent linear control for complex MMAB topologies.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contracts No. APVV-23-0521, No. APVV-19-0210 and No. APVV-18-0436.

References

[1] P. Koohi, A. J. Watson, J. C. Clare, T. B. Soeiro, and P. W. Wheeler, "A survey on multi-active bridge dc-dc converters: Power flow decoupling techniques, applications, and challenges," *Energies*, vol. 16, no. 16, 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/16/5927

[2] Y. Cai, J. Li, C. Gu, J. Yang, S. Guenter, G. Buticchi, and H. Zhang, "A modular modulation decoupling algorithm for multiple active bridge based multiport ev charger," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 12, no. 4, pp. 3637–3649, 2024.

[3] C. Zhao, S. D. Round, and J. W. Kolar, "An isolated three-port bidirectional dc-dc converter with decoupled power flow management," *IEEE Transactions on Power Electronics*, vol. 23, no. 5, pp. 2443–2453, 2008.

[4] A. H. A. Adam, J. Chen, M. Xu, S. Kamel, G. I. Y. Mustafa, Z. A. Zaki, and E. M. Ahmed, "Power decoupling enhancement of a triple active bridge converter with feedforward compensation based on model predictive control and fuzzy logic controller in dc microgrid systems," *IEEE Access*, vol. 12, pp. 140 310–140 328, 2024.

[5] S. Bandyopadhyay, P. Purgat, Z. Qin, and P. Bauer, "A multiactive bridge converter with inherently decoupled power flows," *IEEE Transactions on Power Electronics*, vol. 36, no. 2, pp. 2231–2245, 2021.

[6] P. Purgat, S. Bandyopadhyay, Z. Qin, and P. Bauer, "Power flow decoupling controller for triple active bridge based on fourier decomposition of transformer currents," in *2020 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2020, pp. 1201–1208.

[7] T. Basarik, M. Lacko, A. Grinčová, and K. Kyslan, "Decoupling of mmab converter using the newton iteration method," in *2025 International Conference on Electrical Drives and Power Electronics (EDPE)*, 2025, pp. 1–6.

[8] P. Wang, X. Lu, W. Wang, and D. Xu, "Hardware decoupling and autonomous control of series-resonance-based three-port converters in dc microgrids," *IEEE Transactions on Industry Applications*, vol. 55, no. 4, pp. 3901–3914, 2019.

[9] S. Bandyopadhyay, Z. Qin, and P. Bauer, "Decoupling control of multiactive bridge converters using linear active disturbance rejection," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 10 688–10 698, 2021.

[10] P. Courrieu, "Fast computation of moore-penrose inverse matrices," *arXiv preprint arXiv:0804.4809*, 2008.

Pushing PSFB technology to 500 kHz: What is gained and what is lost?

[1]*Daniel GORDAN (3rd year)*
Supervisor: [2]Marek PÁSTOR

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]daniel.gordan@tuke.sk, [2]marek.pastor@tuke.sk

***Abstract*—This paper compares phase-shifted full-bridge (PSFB) converters operating at 100 kHz and 500 kHz to evaluate the practical effects of increasing switching frequency. While PSFB converters inherently provide high efficiency and natural zero-voltage switching (ZVS), raising the frequency to 500 kHz significantly reduces the size of magnetic components and increases power density. Modern wide-bandgap devices also help limit switching losses at these elevated frequencies. However, higher frequency operation introduces challenges related to reduced inductive energy for maintaining ZVS and stronger influence of parasitic elements. These effects can lead to greater switching losses, more pronounced voltage ringing, and increased stress on the rectifier. The comparison highlights key advantages and limitations at both frequency levels and outlines the trade-offs that must be considered when pushing PSFB converters into the high-frequency regime.**

***Keywords*—PSFB convert, SR rectifier, GaN, high-frequency, ZVS**

I. INTRODUCTION

Power electronic converters have become fundamental elements across a wide spectrum of modern applications, including automotive systems, renewable-energy interfaces, battery storage solutions, and general industrial or consumer electronics. Among isolated high-efficiency topologies, the phase-shifted full-bridge (PSFB) converter has remained one of the most widely used architectures for several decades due to its favorable efficiency, wide input-output operating range, galvanic isolation, and its relatively simple primary-side regulation based on phase-shift control. The combination of natural zero-voltage switching (ZVS) under appropriate conditions, manageable semiconductor stress, and low component count has ensured long-term relevance of the PSFB topology.

Although the fundamental structure of the PSFB converter has not changed since its early implementations [1], significant technological progress has been achieved through the introduction of auxiliary circuits, advanced control approaches, and especially modern wide-bandgap semiconductors. Transistors based on GaN or SiC technology exhibit lower output capacitances, reduced switching and conduction losses, and enable substantially higher switching frequencies [2]-[4]. These properties expand the ZVS operating region and contribute to higher converter efficiency and increased power density. Parallel advancements in microcontroller and DSP platforms make it possible to adaptively optimize timing, dead-time intervals, and switching coordination, further improving soft-switching [5], [6].

Several inherent challenges of PSFB converters, however, remain relevant even with modern semiconductor devices. ZVS strongly depends on load current and becomes increasingly difficult to maintain at light loads [7]. Circulating current contributes to additional conduction losses on the primary side. Different approaches to mitigate these limitations have been explored, including controlled secondary-side rectifiers and various non-dissipative or active snubber networks [7]-[9]. Such auxiliary circuits may also suppress overvoltage ringing on the rectifier, which originates from resonance between transformer leakage inductance and parasitic capacitances.

Recent converter concepts typically combine GaN-based primary inverters with synchronous rectifiers [10]–[12], reducing the number of secondary-side switches while maintaining ZVS and high efficiency over a broad operational range. These improvements simplify the converter structure and further support operation at elevated switching frequencies.

As switching frequency increases, the potential for reducing transformer and inductor size becomes one of the most important motivations for pushing the PSFB converter beyond conventionally used frequencies near 100 kHz. The transition toward several-hundred-kilohertz operation, such as 500 kHz, can significantly increase power density due to reduced magnetic volume. However, higher frequencies also intensify challenges associated with ZVS conditions, circulating current, transformer design, semiconductor output capacitance, and rectifier overvoltage behaviour. Understanding these trade-offs is therefore essential for evaluating the realistic benefits and limitations of operating a PSFB converter at substantially higher switching frequencies.

II. PSFB CONVERTER

The phase-shifted full-bridge (PSFB) converter, Fig. 1, consists of a full-bridge inverter on the primary side, a high-frequency transformer providing galvanic isolation, and a rectifier followed by an output filter on the secondary side. The inverter is formed by four semiconductor switches arranged in a full-bridge configuration. Each switch operates with a duty ratio of 50%, while the effective power transfer is controlled by adjusting the phase shift between the two bridge legs.

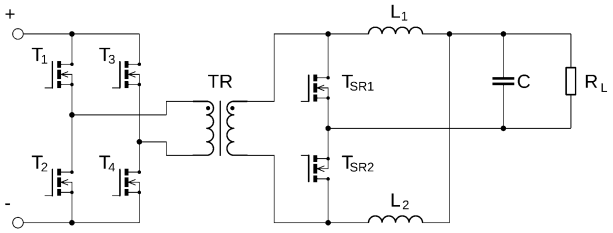


Fig. 1. The schematic of the PSFB converter with current doubler rectifier

The transformer transfers this high-frequency square-wave voltage to the secondary side, where it is processed by the selected rectifier structure. Depending on the application, the rectifier may be implemented as a full-bridge, centre-tapped, or current-doubler configuration, each with its own trade-offs in terms of conduction losses, transformer design, and current distribution. In this case a current doubler rectifier (CD) synchronous rectifier is used due to the benefits at high output current.

ZVS is necessary to achieve low switching losses, especially at high switching frequencies. The condition for achieving ZVS depends on the amount of inductive energy in a series resonant circuit formed by the leakage inductance and parasitic output capacitance of the inverter transistor. For the output-capacitance voltage to fall to zero before the transistor is turned on, the inductive energy available in the resonant circuit must exceed the capacitive energy. Inductive energy is different for transistor T_1 a T_2 (leading leg) and T_3 a T_4 (lagging leg):

$$E_{L_{leading}} = 0{,}5Lk\left(I_{M(pk)} + I_{L1(min)}\frac{N_S}{N_P}\right)^2, \quad (1)$$

$$E_{L_{lagging}} = 0{,}5L_M I_{M(pk)}^2 + 0{,}5L_1\left(\frac{N_P}{N_S}\right)^2\left(I_{L1(\max)}\frac{N_S}{N_P}\right)^2 + 0{,}5Lk\left(I_{M(pk)} + I_{L1(max)}\frac{N_S}{N_P}\right)^2, \quad (2)$$

so that the leading leg achieves ZVS only over a smaller range than the lagging leg.

III. Laboratory Measurements

This section discusses the differences between 100 kHz and 500 kHz PSFB converters operating with the same parameters, as is shown in Table. 1. A photo of the 500 kHz converter is shown in Fig. 2. The inverter is identical in both prototypes, while the 100 kHz version differs only in the transformer and rectifier. The parameters of the transformers are listed in Table 2, together with the output filter parameters. The transformer designed for higher-frequency is planar, providing substantially lower leakage inductance. Although the magnetizing inductance remains comparable to that of the 100 kHz design, the resulting magnetizing current at 500 kHz is considerably smaller. Consequently, based on equations (1) and (2), achieving ZVS with the same inverter transistor (same output capacity) becomes more challenging, particularly for the leading leg. Lower leakage inductance causes faster current changes, which affect the occurrence of oscillations and the reliable operation of the converter, and this can also be seen in Fig. 3 (waveforms of the 100 kHz converter) and Fig. 4 (waveforms of the 500 kHz converter).

TABLE I
Parameters of PSFB converters

Parameter	Value
Supply voltage V_{IN}	390 V
Output current I_O	25 A
Output voltage V_O	24 V
Output power P	600 W
Switching frequency fs	100 kHz or 500 kHz
Inverter transistors	GS-065-011-2
Rectifier transistors (100 kHz)	IPTG039N15NM5
Rectifier transistors (500 kHz)	IGC037S12S1

TABLE II
Transformer and output filter parameters

Parameter	100 kHz	500 kHz
transformer volume (cm^3)	145,8	**29,5**
magnetizing inductance(µH)	1013	726,9
leakage inductance (µH)	13,29	3,57
filter inductance (µH)	60	**12**
filter capacitance (µF)	56	**10**

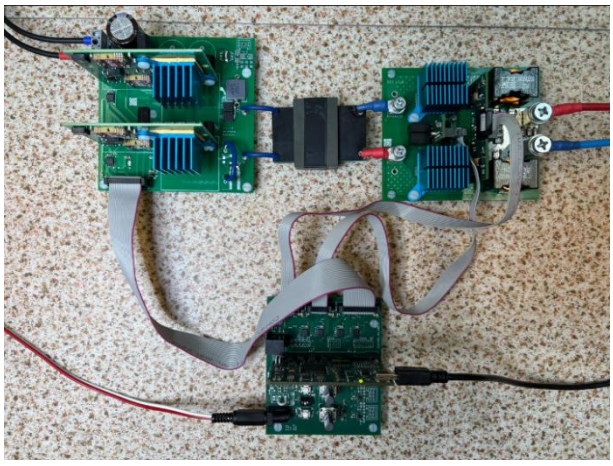


Fig. 2. The 500 kHz PSFB converter

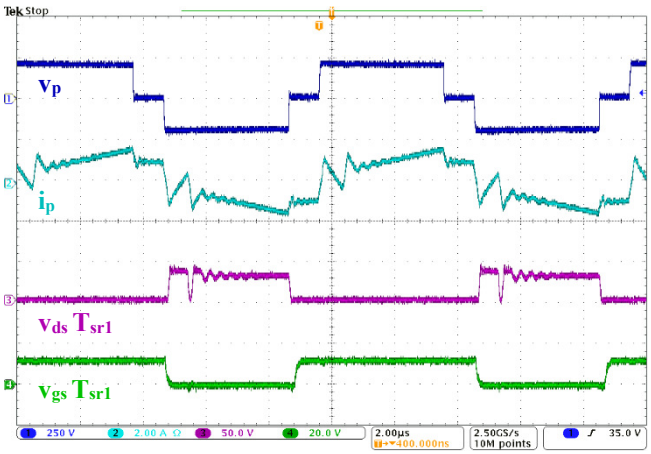


Fig.3. The waveforms of 100 kHz PSFB converter at half output load

It is possible to observe incomplete ZVS in the 500 kHz converter at half load in the leading leg, indicated by the overvoltage-related oscillation of the primary voltage. Due to the higher slew rates and increased sensitivity to parasitic elements, additional oscillations appear, which are visible not only on the primary side—both in the primary current and voltage—but also in the voltage of the synchronous rectifier.

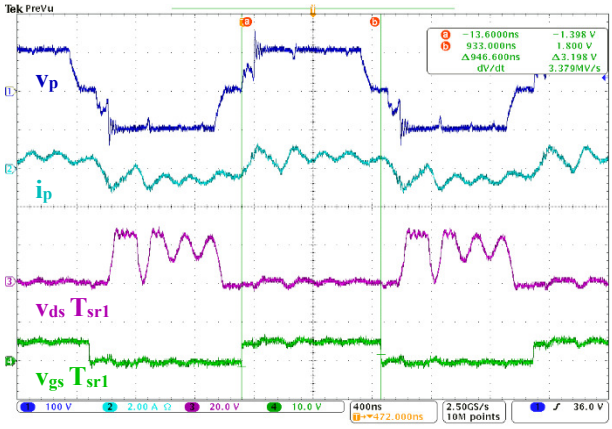


Fig. 4. The waveforms of 500 kHz PSFB converter at half output load

By redesigning the transformer to reduce the magnetizing inductance, it was possible to increase the magnetizing current at 500 kHz and achieve full ZVS at 50% load, Fig. 5. Further design details will be presented in the forthcoming publication. As can be seen from the waveform, the primary-side voltage no longer exhibits overvoltage-related ringing, while the primary current and the rectifier voltages still show slight oscillations. The efficiency comparison is shown in Fig. 6. Based on the waveforms, it can be seen that the 100 kHz converter achieved higher efficiency, which is primarily due to its lower switching losses—strongly dependent on the switching frequency. A useful design guideline for PSFB converters is to ensure full ZVS for the leading leg from approximately 50% load, which is also indirectly reflected in the measured waveforms, as both converters reached their maximum efficiency near this load level.

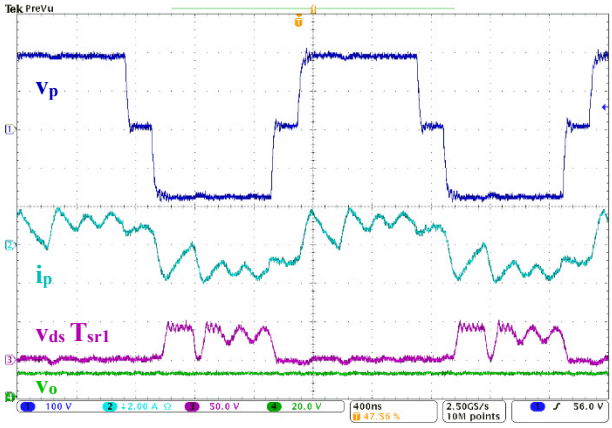


Fig. 5. The waveforms of 500 kHz PSFB converter at half output load with increased magnetizing current

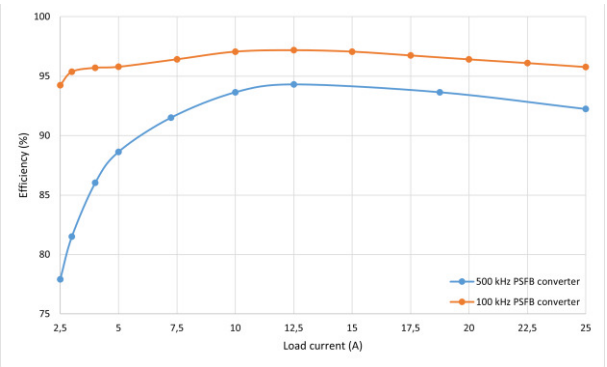


Fig. 6. Comparison of efficiency 100 kHz and 500 kHz PSFB converter

IV. Conclusion

The PSFB converter is a proven topology, and efforts to improve its efficiency have continued throughout its development. Before the introduction of modern GaN and SiC devices, especially when using IGBTs, achieving high efficiency even at 50 kHz required careful optimization together with auxiliary networks [13]. Wide-bandgap transistors later enabled the removal of these additional circuits, whose parasitics were disadvantageous, and made it possible to reach good efficiency at around 100 kHz [14]. Even at these frequencies, proper coordination between the transformer design and the devices output capacitances remains essential [15]. When the switching frequency is increased toward 500 kHz, several earlier issues re-emerge, most notably the difficulty of achieving ZVS in the leading leg. This is reflected in the noticeably lower efficiency below approximately 25% load. Moreover, higher slew rates and sensitivity to parasitics result in increased ringing. On the positive side, increasing the frequency significantly reduces the size of the transformer, inductors, and capacitors, leading to higher power density. Therefore, future work should focus on strategies to widen the ZVS range and improve efficiency. Transformer optimization-briefly demonstrated in this paper-has already shown that proper design can restore ZVS, improve efficiency, and suppress oscillations. As frequency rises, power density improves, but so do the design demands. Thus, 500 kHz should be viewed not as an endpoint, but as another milestone toward further development.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under the project APVV-23-0521 and VEGA 1/0584/24.

References

[1] L. H. Mweene, C. A. Wright, M. F. Schlecht, “A 1 kW, 500 kHz frontend converter for a distributed power supply system,” APEC, 1989.
[2] M. Beheshti, “Wide-bandgap semiconductors: Performance and benefits of GaN versus SiC,” Analog Design Journal, TI, 2020.
[3] M. Heintze, I. S. Butzmann, “A GaN 500 kHz High Current Active Clamp Phase-Shifted Full-Bridge Converter…,” EPE’18 ECCE Europe, 2018.
[4] F. Qi, Z. Wang, Y. Wu, P. Zuk, “Advantage of GaN in Phase Shift Full Bridges,” WiPDA, 2019.
[5] Texas Instruments, Phase-Shifted Full Bridge DC/DC Power Converter Design Guide, 2014.
[6] H. Liu et al., “A Novel PID Control Strategy… for PSFB Current-Doubler SR Converter,” IMCEC, 2021.
[7] Y.-D. Kim et al., “Wide-Range ZVS Phase-Shift Full-Bridge Converter With Reduced Conduction Loss…,” IEEE TPEL, 2013.
[8] M. Pastor, J. Dudrik, R. Michal, “High-Frequency Soft-Switching DC-DC Converter with Simple Secondary Turn-Off Snubber,” PEMC, 2022.
[9] S.-Y. Lin, C.-L. Chen, “Analysis and design for RCD clamped snubber…,” IEEE TIE, 1998.
[10] M. Heintze, S. Butzmann, “A GaN 500 kHz High Current Active Clamp Phase-Shifted Full-Bridge…,” EPE'18.
[11] Infineon Technologies AG, “1000 W 400 V phase-shifted full-bridge current doubler…,” 2020.
[12] Texas Instruments, “3-kW Phase-Shifted Full Bridge With Active clamp
[13] M. Pástor, D. Gordan, T. Basarik, C.L. Maxwell, "Inverter optimization of soft-switching DC-DC converter," in Automatika, vol. 65, no. 4, pp. 1391–1401, 2024, https://doi.org/10.1080/00051144.2024.2388442
[14] M. Pástor, D. Gordan, C.L. Maxwell, From IGBT to GaN implementation in phase-shifted full-bridge converter. Electr Eng 107, 11831–11845 (2025). https://doi.org/10.1007/s00202-025-03125-4
[15] D. Gordan, M. Pástor and T. Kmecik, "Comparative Study of PSFB Converter Efficiency for Various GaN Semiconductor Devices," in Acta Polytechnica Hungarica, vol. 22, no. 11, pp. 27-44, 2025

Multi-Agent Reinforcement Learning in Energy Markets: State and Outlook

[1]*Michaela LEL'OVÁ (1st year),*
Supervisor: [2]*Ivana BUDINSKÁ, Consultant:* [3]*Erik KAJÁTI*

[1,3]Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovak Republic
[2]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovak Republic

[1]michaela.lelova@tuke.sk, [2]ivana.budinska@savba.sk, [3]erik.kajati@tuke.sk

***Abstract*—This paper proposes a reproducible multi-agent reinforcement learning (MARL) framework for realistic energy-market environments, augmented with structured explainable AI mechanisms. The environment is built from open European power-system data, enabling large-scale simulations that reflect renewable variability, demand uncertainty, and price dynamics. Multiple agents—such as battery operators and generators—learn charging, dispatch, and bidding strategies in cooperative and competitive settings using state-of-the-art MARL algorithms. The framework evaluates convergence, scalability, equilibrium behaviour, and robustness under scenario shifts, explicitly addressing non-stationarity and increasing system complexity. An integrated explainability layer provides global and local policy insights through structured abstraction and query-based methods, improving transparency and interpretability. The resulting framework serves as a benchmark for studying scalable and interpretable MARL in strategic energy-market applications.**

***Keywords*—Industry 5.0, decision-support framework, multi-agent reinforcement learning, explainable ai**

I. Introduction

This paper proposes the design and evaluation of a multi-agent reinforcement learning (MARL) framework for complex energy-market environments, augmented with explainable artificial intelligence (XAI) mechanisms. The goal is to study learning dynamics, coordination, and interpretability in a realistic large-scale multi-agent setting constructed from open European power-system data and models.

Recent work has demonstrated the effectiveness of MARL in decentralized control problems such as microgrids, industrial energy systems, and distribution networks. However, many of these studies focus primarily on application-level performance and do not deeply investigate replicability and interpretability in strategic multi-agent environments. Furthermore, most energy-domain MARL research operates at the device-control level, whereas strategic interaction among learning agents in market-like environments remains comparatively under explored from a machine learning perspective.

The core research idea of this paper is to study how learning agents—such as battery operators, electricity producers, and flexible demand participants—interact in a simplified but realistic power-market environment and learn charging, bidding, and dispatch strategies. The environment uses open European datasets and established power-system modelling frameworks (e.g., PyPSA and ENTSO-E fundamentals) to reflect real variability in renewables, demand, and prices. We further integrate explainability methods into this MARL engine to make agent decisions more transparent for human analysts [1]. This addresses a key gap, as many MARL-based energy optimization models remain difficult to justify to risk managers, regulators, and market participants.

The work is motivated by Europe's current combination of high energy prices, security risks, and decarbonisation pressure, which increases the need to optimize flexible assets and trading strategies [2]. Although market participants already use AI for forecasting and optimization, many models are difficult to justify to risk managers, regulators, and clients. Our goal is to show that MARL can be integrated into a human-centric AI workflow aligned with Industry 5.0 principles of transparency, resilience, and sustainability.

II. Literature Review and Analysis

This section reviews the application of MARL in microgrid and macrogrid environments.

A. The use of MARL in microgrid energy systems

Recent studies have shown that the use of hierarchical reinforcement learning, and in particular the options framework, to address typical energy scheduling problems in microgrids leads to lower operational costs, improved generalization, faster convergence, and enhanced training stability when compared to primitive (flat) reinforcement learning approaches [3].

Among the evaluated MARL methods, multi-agent option-critic deep deterministic policy gradient (MA-OCDDPG) outperformed Multi-Agent Deep Deterministic Policy Gradient (MADDPG) and Multi-Agent Advantage Actor-Critic (MAA2C). Its advantage comes from combining hierarchical reinforcement learning's structured decisions with MARL coordination. It also achieved performance comparable to model-based rolling-horizon optimization under variable renewable generation and uncertain demand, while preserving the flexibility of data-driven adaptive control.

Notably, the approach scales effectively to larger systems with an increasing number of distributed energy resources, operates without requiring real-time inter-agent communication, satisfies load constraints, reduces energy costs, and increases renewable energy utilization [3]. These characteristics make it particularly well suited for large-scale microgrid deployments.

Despite these promising results, several limitations remain. The existing formulation is restricted to fully cooperative

scenarios, and its extension to semi-cooperative or competitive multi-agent settings has not yet been explored. In addition, as system complexity increases, the selection and tuning of hyperparameters become more challenging, potentially limiting practical scalability and robustness.

A different approach is presented in [4], which, similar to [3], aims to improve upon the standard MADDPG framework. The authors propose LSD-MADDPG (Local Strategy-Driven MADDPG) for energy management systems, demonstrating strong performance in both cooperative and competitive settings. In this formulation, agents represent individual energy consumers.

Traditional centralized reinforcement learning methods—such as Q-learning [5], Deep Q-Networks (DQN) [6], Double DQN [7], and Deep Deterministic Policy Gradient (DDPG) [8]—are widely used in energy management. However, they raise privacy concerns due to centralized data collection, remain vulnerable to single-point failures, and scale poorly as system size increases.

Decentralized and multi-agent reinforcement learning methods address some of these issues by enabling parallel computation and improving system reliability through agent redundancy. Nevertheless, typical MARL approaches often struggle to achieve objectives that require tight coordination among agents and frequently underperform when compared to centralized strategies.

To mitigate these limitations, [4] adopts the centralized training with decentralized execution (CTDE) paradigm. The primary objectives of the proposed approach are to flatten the aggregate demand curve, reduce overall energy consumption, and maintain customer satisfaction.

In LSD-MADDPG, agents do not share a centralized critic; instead, they exchange only local energy consumption strategies. Despite this reduced information sharing, LSD-MADDPG achieves performance comparable to CTDE-MADDPG while offering improved scalability, making it well suited for large-scale energy management systems [4].

[9] proposes a two-layer reinforcement learning optimization framework for the shift in distribution networks from unidirectional operation to bidirectional, highly interactive power flow. In this setting, the grid faces substantial uncertainty from renewable generation fluctuations, load variability, external disturbances, and increasing dynamic loads. High renewable penetration further increases system stochasticity, making multi-objective optimization and real-time control more difficult.

To address these challenges, the authors use hierarchical reinforcement learning. Upper-layer agents handle global decisions, including voltage optimization, fault-area identification, and recovery policy design. They target long-term objectives such as network-wide power-flow optimization, more economical equipment scheduling, and lower operating costs.

Lower-layer agents operate at shorter time scales for device-level scheduling and control. They execute fast actions at roughly 15-minute resolution, enabling immediate component-level responses to maintain short-term voltage stability. This hierarchical decomposition coordinates long-term strategic planning with rapid local control in uncertain distribution-network environments.

Another relevant framework is proposed in [10], which focuses on the aforementioned uncertainty inherent in renewable energy sources, by introducing a multi-agent deep reinforcement learning (MADRL) approach for energy management in flexible switchboards (SBs). In this setting, the switchboard enables flexible energy distribution by supplying loads at multiple voltage levels and integrating photovoltaic (PV) generation and energy storage systems across multiple buses.

The proposed method decomposes the SB into multiple energy local area networks (E-LANs), which may operate in either cooperative or competitive modes. Each E-LAN is managed by a MADDPG agent responsible for optimizing local energy distribution, while interactions among E-LANs enable system-wide coordination. Despite decentralized decision-making, the framework achieves up to 11.5% cost savings compared to baseline strategies [10].

Uncertainty in PV generation is addressed through a scenario-based sampling approach, whereby probabilistic (chance) constraints are transformed into deterministic constraints using forecast samples. This enables a model-free solution via MADDPG, avoiding the need for explicit stochastic modelling of renewable uncertainty.

Simulation results demonstrate that the proposed approach significantly reduces operating costs compared to single-policy (SP) and single-agent DDPG methods. Furthermore, energy trading among E-LANs introduces additional cost savings by leveraging both cooperative and competitive interactions within the switchboard architecture [10].

An industrially oriented MARL framework is presented in [11] and validated in a real-world steel powder manufacturing facility. Unlike most existing studies, which primarily focus on residential and commercial buildings, this work addresses the application of MARL in heavy industry, thereby highlighting an important research gap. The authors emphasize that standardization efforts and benchmark environments in energy-related reinforcement learning research are still largely centred on small-scale building applications, leaving industrial use cases under explored.

The energy management problem is formulated as a partially observable Markov decision process to capture uncertainties in energy consumption and production processes. The training procedure follows a CTDE paradigm and integrates an attention mechanism to enhance coordination and policy learning among agents.

Despite its contributions, the framework has several limitations. The current formulation is restricted to price-based demand response, where electricity price signals serve as the primary mechanism for load scheduling. Renewable energy integration and broader flexibility mechanisms are not fully incorporated. Future work could explore the simultaneous application of multiple DR schemes to enable more complex and adaptive energy management strategies.

Moreover, the optimization problem is modelled in discrete time, where load interruptions are treated as discrete events. Extending the framework to continuous control formulations could enable smoother and more realistic operational policies. There is also significant potential to expand the DR framework beyond electricity to include additional utilities such as heat, hydrogen production, and other industrial energy carriers.

However, increasing system scope and complexity would substantially raise computational demands. Future research directions therefore include IIoT-integrated MARL architectures and system-level design, parallelized training and hardware acceleration, model compression for edge deployment,

transitioning from day-ahead scheduling to near real-time control, and further algorithmic extensions with standardized benchmarking [11].

Overall, several fundamental challenges remain for MARL in microgrid and distribution-network control. A primary issue is non-stationarity: concurrent policy updates continuously change transition dynamics and reward functions, which makes convergence difficult. This issue is especially severe in dynamic tasks such as frequency and voltage regulation. Renewable intermittency further worsens the problem because agents must adapt to highly variable generation without stable training data [12].

Another major limitation is the insufficient incorporation of physical and safety constraints. Many existing MARL formulations fail to adequately model grid operational constraints, such as voltage limits, line capacities, and protection requirements, raising concerns about safety and real-world deployability [13], [9], [4].

Multiple tightly coupled time scales add further complexity. Long-term planning objectives (e.g., economic dispatch and equipment scheduling) and short-term control actions (e.g., voltage and frequency regulation) are interdependent, because decisions at one scale affect system behaviour at others. This coupling rapidly expands the action space and motivates layered, distributed control architectures in which local agents act autonomously while coordinating to preserve global grid stability and efficiency [9].

Finally, existing approaches often exhibit limited robustness, scalability, generalization capability, and risk sensitivity, particularly under unseen operating conditions or extreme events [12], [9]. As the number of agents and controllable devices increases, system complexity grows exponentially, further hindering training efficiency and practical scalability [3], [13], [11].

B. The use of MARL in macrogrid energy systems

Both [14] and [15] extend traditional MARL to handle complex and uncertain systems, but they follow different approaches. QiMARL [14] introduces a quantum-inspired mechanism to improve exploration in large and high-dimensional state-action spaces. By using ideas from quantum parallelism, it can evaluate multiple possible solutions simultaneously, which is particularly useful for resource distribution in nodal power stations. Its main advantage lies in managing uncertainty and combinatorial complexity more efficiently than standard MARL methods.

In contrast, the MARL–VI framework proposed in [15] combines MARL with variational inequality (VI) theory to guide agents toward stable equilibrium solutions. Instead of focusing mainly on improving exploration, this approach emphasizes stability, convergence guarantees, and faster training. The hybrid framework significantly reduces training time compared to baseline MARL algorithms such as MAPPO and MADDPG, making it suitable for large-scale and time-sensitive applications. So while QiMARL mainly improves how agents explore and search in complex environments, the MARL–VI framework improves convergence speed and theoretical stability.

C. Explainability in MARL

Explainability in MARL remains an emerging research direction, with recent works proposing structured methods to make complex multi-agent policies more interpretable.

In [16], the authors propose several techniques for generating policy explanations in MARL systems and evaluate them in domains such as multi-robot warehouse coordination and search-and-rescue tasks. Their framework is based on modelling the system as a MMDP, enabling both global and local explanations without requiring access to the internal structure of the learning algorithm.

The framework introduces two explanation mechanisms. First, *policy abstraction and summarization* provide a global view of learned behaviour. When policies become complex because of large state spaces or many interacting agents, an abstract MMDP is built during evaluation, summarized as a weighted graph, and analyzed through its most probable execution paths. Second, *query-based explanations* provide local, user-specific insights. Following earlier single-agent work (e.g., [17]), the framework supports *when*, *what*, and *why not* queries. It converts the MMDP into a Boolean formula, minimizes it with the Quine–McCluskey algorithm, and translates the result into natural-language explanations. To improve interpretability, relevance filtering (WithRF) removes irrelevant agents and state features, producing simpler explanations than the non-filtered (NoRF) approach.

Building on this work, [18] extends the framework to temporal queries in MARL systems. The approach first verifies whether a user query is feasible before generating an explanation, improving reliability and consistency. Although demonstrated with one MARL algorithm, the authors argue it should generalize to other methods and identify cross-algorithm validation and more complex user queries as key future directions. They also emphasize testing in additional domains and environments to assess scalability and robustness.

A broader view of explainable reinforcement learning is provided by [19]. The survey groups methods into two categories: approaches that are interpretable by design and approaches that explain or repair already trained, non-interpretable models. It also examines causal claims in explanation methods, highlighting that many provide correlational rather than truly causal insight. In addition, the authors distinguish methods that introduce new explanation types from those that evaluate or refine existing ones. Their taxonomy further separates type-based approaches (optimizing one explanation form) from objective-based approaches (optimizing explanation quality across forms). Overall, the survey offers practical guidance for selecting explainability methods based on reinforcement learning system characteristics and requirements.

III. Problem Identification and Future Directions

Future work will follow a comparative research program rather than a single optimization target. We will assess which AI paradigms are most suitable for multi-agent energy coordination by comparing MARL with centralized optimization and forecasting-control pipelines, rule-based and game-theoretic strategies, and selected statistical methods used in power-system operation. The objective is not to identify one universally optimal strategy, but to provide evidence-based guidance on where each method class is most appropriate.

The empirical focus will be grid congestion and EV-flexibility management across multiple system scales. We begin with microgrid scenarios and progressively extend to macrogrid and market-participant settings, including charging,

dispatch, and congestion-mitigation tasks under realistic demand and renewable variability. This enables a structured assessment of transferability: which methods developed for local energy communities can be generalized to larger transmission-level or market-coupled environments, and under what assumptions.

Methodologically, we will define a common evaluation protocol and comparison tables across tasks and algorithms. Evaluation criteria will include efficiency (cost, peak reduction, load balancing), stability (training convergence, policy consistency), scalability (number of agents), robustness (demand spikes, renewable intermittency), and practical deployability (data requirements, computational complexity, interpretability). We will compare centralized versus decentralized control, cooperative versus competitive interaction, communication versus no-communication settings, and learning-based versus rule-based agent behaviour.

At least one representative CTDE MARL baseline (e.g., MADDPG or MAPPO) will be implemented together with selected non-MARL baselines to ensure scientifically meaningful comparisons [15]. Building on this, we will develop an architecture proposal for a reproducible multi-agent simulation environment that can later be instantiated in a physical proof-of-concept platform (e.g., small scale physical model). This step connects theoretical modelling choices with practical implementation constraints.

Finally, we will treat explainability as a design requirement. Following prior MARL explanation work [16], [18], [19], we will integrate structured explanation tools into the benchmarking workflow so that model comparisons include not only performance metrics but also transparency and decision traceability. The intended doctoral contribution is a generalizable modelling and evaluation procedure for multi-agent energy management, with practical recommendations for method selection across grid sizes, demand patterns, and pricing mechanisms.

IV. Conclusion

This doctoral research proposes a comparative framework for AI-based multi-agent energy management, with emphasis on grid congestion and EV flexibility across microgrid and macrogrid settings. We will benchmark MARL and non-MARL approaches using a unified set of criteria (efficiency, stability, scalability, robustness, deployability, and explainability) to identify where each strategy is most suitable. The expected outcome is a generalizable modelling and evaluation procedure with practical recommendations for real-world implementation.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences VEGA 2/0135/23.

References

[1] G. P. Reddy and Y. V. Kumar, "Explainable ai (xai): Explained," *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences, eStream 2023 - Proceedings*, 2023.

[2] E. Commission, D.-G. for Research, and Innovation, *Enabling Technologies for Industry 5.0 – Results of a workshop with Europe's technology leaders*. Publications Office, 2020.

[3] I. Jendoubi and F. Bouffard, "Multi-agent hierarchical reinforcement learning for energy management," *Applied Energy*, vol. 332, p. 120500, 2 2023.

[4] P. Wilk, N. Wang, and J. Li, "Multi-agent reinforcement learning for smart community energy management," *Energies 2024, Vol. 17, Page 5211*, vol. 17, p. 5211, 10 2024. [Online]. Available: https://www.mdpi.com/1996-1073/17/20/5211/htmhttps://www.mdpi.com/1996-1073/17/20/5211

[5] S.-J. Chen, W.-Y. Chiu, and W.-J. Liu, "User preference-based demand response for smart home energy management using multiobjective reinforcement learning," *IEEE Access*, vol. 9, pp. 161 627–161 637, 2021.

[6] A. Mathew, A. Roy, and J. Mathew, "Intelligent residential energy management system using deep reinforcement learning," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5362–5372, 2020.

[7] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 572–582, 2020.

[8] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2751–2762, 2019.

[9] N. Xu, Z. Tang, C. Si, J. Bian, and C. Mu, "A review of smart grid evolution and reinforcement learning: Applications, challenges and future directions," *Energies 2025, Vol. 18, Page 1837*, vol. 18, p. 1837, 4 2025. [Online]. Available: https://www.mdpi.com/1996-1073/18/7/1837/htmhttps://www.mdpi.com/1996-1073/18/7/1837

[10] J. Deng, X. Wang, and F. Meng, "Multi-agent deep reinforcement learning for smart building energy management with chance constraints," *Energy and Buildings*, vol. 331, p. 115408, 3 2025.

[11] A. Bashyal, T. Boroukhian, P. Veerachanchai, M. Naransukh, and H. Wicaksono, "Multi-agent deep reinforcement learning based demand response and energy management for heavy industries with discrete manufacturing systems," *Applied Energy*, vol. 392, p. 125990, 8 2025.

[12] S. Keren, C. Essayeh, S. V. Albrecht, and T. Morstyn, "Multi-agent reinforcement learning for energy networks: Computational challenges, progress and open problems," 5 2024. [Online]. Available: http://arxiv.org/abs/2404.15583

[13] P. Kofinas, A. Dounis, and G. Vouros, "Fuzzy q-learning for multi-agent decentralized energy management in microgrids," *Applied Energy*, vol. 219, pp. 53–67, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918303465

[14] S. M. Turjya, A. Bandyopadhyay, M. S. Kaiser, and K. Ray, "Qimarl: Quantum-inspired multi-agent reinforcement learning strategy for efficient resource energy distribution in nodal power stations," *AI 2025, Vol. 6,*, vol. 6, 9 2025. [Online]. Available: https://www.mdpi.com/2673-2688/6/9/209

[15] L. D. Natale, G. Fargetta, L. R. Scrimali, and S. Battiato, "Multi-agent reinforcement learning and variational inequality models for international trade networks under crisis," *Simulation Modelling Practice and Theory*, vol. 146, p. 103219, 1 2026.

[16] K. Boggess, S. Kraus, and L. Feng, "Toward policy explanations for multi-agent reinforcement learning," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 109–115, 4 2022. [Online]. Available: https://arxiv.org/abs/2204.12568v4

[17] B. Hayes and J. A. Shah, "Improving robot controller transparency through autonomous policy explanation," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. Part F127194, pp. 303–312, 3 2017.

[18] K. Boggess, L. Feng, and S. Kraus, "Explainable multi-agent reinforcement learning for temporal queries," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2023-August, pp. 55–63, 5 2023. [Online]. Available: https://arxiv.org/abs/2305.10378v1

[19] S. Milani, N. Topin, M. Veloso, and F. Fang, "Explainable reinforcement learning: A survey and comparative review," *ACM Computing Surveys*, vol. 56, pp. 1–36, 7 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3616864

Bidirectional DC/DC Converters: An Overview

[1]*Roland MOLNÁR (1st year),*
Supervisor: [2]*Marek PÁSTOR*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Košice, Slovak Republic

[1]roland.molnar@tuke.sk, [2]marek.pastor@tuke.sk

***Abstract*—Bidirectional DC–DC converters have become a key enabling technology in modern energy systems, including electric vehicles, energy storage systems, and DC microgrids. Their ability to provide controlled bidirectional power flow, high efficiency, and wide voltage conversion capability makes them essential for emerging applications such as fast EV charging and Vehicle-to-Grid operation. This paper presents an overview of bidirectional DC–DC converter topologies with a focus on isolated solutions suitable for high-power applications. Particular attention is given to the Dual Active Bridge converter, advanced phase-shift modulation techniques, and modern control strategies. Current application trends and system-level aspects are discussed, and key research challenges related to efficiency optimization, control complexity, and wide operating voltage ranges are identified. The presented analysis highlights current design directions and outlines future research needs for high-performance bidirectional power conversion systems.**

***Keywords*—Bidirectional DC–DC converter, Dual Active Bridge, Electric Vehicles, Isolated Converters**

I. Introduction

The global transition toward environmental sustainability and the rapid electrification of transportation systems have significantly increased the demand for renewable energy sources and advanced power conversion technologies [1], [2]. In this context, bidirectional DC–DC converters (BDCs) have become key components of modern energy systems, as they enable controlled power flow between energy storage systems (ESS), such as lithium-ion batteries or supercapacitors, and DC or AC grids [3]. Their capability to operate in both buck and boost modes makes them essential for applications including smart grids and DC microgrids.

Recent research has focused on achieving high power density and efficiency to support demanding applications such as electric vehicle (EV) fast charging and Vehicle-to-Grid (V2G) operation [4], [5]. This trend has accelerated the adoption of Wide Bandgap (WBG) semiconductor devices, particularly Gallium Nitride (GaN) and Silicon Carbide (SiC), which enable higher switching frequencies and allow a significant reduction of magnetic component size [1]. However, higher switching frequencies and wide operating voltage ranges introduce new challenges related to system stability, efficiency optimization, and dynamic performance under variable load conditions [5].

II. Classification of Bidirectional DC/DC Converters

To capture the diversity of bidirectional DC–DC converters (BDCs), a clear classification based on technical and functional requirements is required [3]. This classification provides a practical framework for selecting a suitable converter topology for a given power system and application context [3], [6].

In the literature, BDCs are commonly classified according to several key criteria, including galvanic isolation, power rating, voltage conversion range, and application domain. The most fundamental distinction is based on galvanic isolation, which divides converters into non-isolated (NBDC) and isolated (IBDC) types [3], [7]. While NBDCs operate without a transformer, IBDCs employ a high-frequency transformer (HFT) to provide electrical isolation and enable significantly higher voltage conversion ratios. Further classification considers power levels ranging from low-power portable systems to high-power industrial and EV fast-charging applications, as well as the ability to interface low-voltage energy storage units (e.g., 48 V batteries) with high-voltage DC buses [1], [5]. In addition, application requirements in energy storage systems, electric vehicles, and DC microgrids impose different constraints on efficiency, power density, and dynamic performance [8], [9].

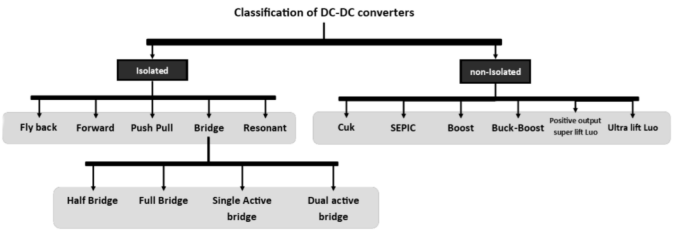


Fig. 1. Classification of DC–DC converters [7]

A. Non-Isolated vs. Isolated Converters

Non-isolated converters are characterized by a simple structure, low component count, and cost-effectiveness, making them attractive for low-voltage and medium-power applications [7], [10]. Classical topologies such as buck–boost, SEPIC, and Cuk converters offer high efficiency under moderate conversion ratios but suffer from increased switch stress and efficiency degradation when extreme duty cycles are required for high voltage gain [10]. Moreover, the absence of galvanic isolation limits their use in high-voltage environments due to safety concerns and increased susceptibility to electromagnetic interference (EMI).

In contrast, isolated converters overcome these limitations by incorporating an HFT, which ensures electrical isolation and improves operational safety [3], [7]. Topologies such as the Dual Active Bridge (DAB) and LLC (inductor–inductor–capacitor) resonant converters are widely adopted in high-power applications, as they enable soft-switching operation (ZVS/ZCS) over a wide load range and support high

power density designs [5], [10]. Although isolated converters are inherently more complex and bulky due to the transformer, their superior efficiency, scalability, and compliance with safety standards have established them as the preferred solution for EV charging systems and DC microgrid integration [3].

III. Non-Isolated DC/DC Converters

Non-isolated bidirectional DC–DC converters (NBDCs) are commonly used in applications where low cost, reduced volume, and high efficiency are prioritized and galvanic isolation is not strictly required [3], [8]. These converters are typically derived from classical unidirectional topologies by replacing unidirectional switches with bidirectional (two-quadrant) switches, enabling power flow in both directions [3].

The most basic NBDC implementation consists of a buck and boost converter connected in an anti-parallel configuration, as illustrated in Fig. 2. Because of their simple structure and low component count, buck–boost-based converters are widely adopted in low-voltage and low-power applications, such as battery-powered devices and automotive auxiliary systems. [3]

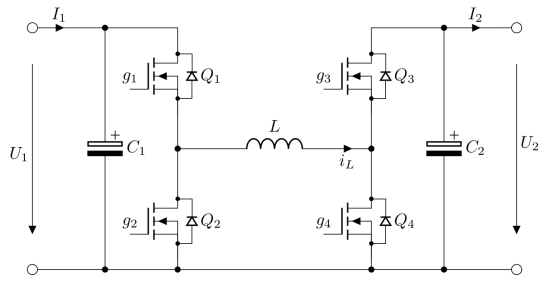


Fig. 2. Bidirectional DC–DC Buck-Boost Bridge Converter

However, their voltage conversion capability is inherently limited, typically below a 4:1 ratio [7]. Achieving higher voltage gain requires extreme duty cycles, which results in increased semiconductor stress, higher conduction losses, and reduced overall efficiency [11]. In addition, the absence of galvanic isolation restricts their use in high-voltage DC environments due to safety concerns and sensitivity to input disturbances [8].

To extend the power capability of non-isolated converters, interleaved structures are frequently employed, particularly in EV and DC microgrid applications [3]. By operating multiple converter phases in parallel with appropriate phase shifting, interleaving significantly reduces input and output current ripple, lowers current stress on individual switches, and improves power density [12]. These advantages come at the expense of increased component count and more complex control requirements, which limit the scalability of NBDCs in high-power and high-voltage systems [8].

IV. Isolated Bidirectional DC/DC Converters

Isolated bidirectional DC–DC converters (IBDCs) represent the preferred solution for high-power applications such as grid-connected energy storage systems and electric vehicle (EV) charging, where safety, reliability, and wide voltage conversion capability are critical. By employing a high-frequency transformer (HFT), these topologies (Fig. 3, Fig. 4) provide galvanic isolation, mitigate common-ground related electromagnetic interference (EMI), and enable significantly higher voltage conversion ratios compared to non-isolated converters [3], [7]. As a result, IBDCs dominate modern EV charging infrastructures and DC microgrid interfaces.

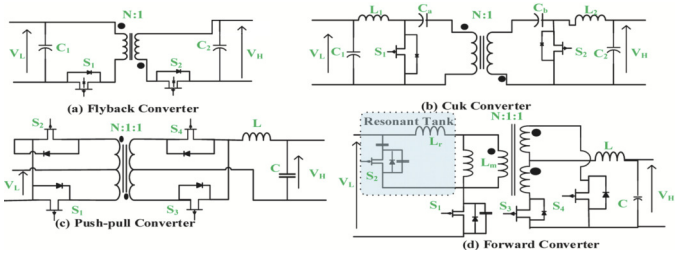


Fig. 3. Examples of isolated bidirectional DC–DC converters [3]

Among isolated topologies, the Dual Active Bridge (DAB) converter (Fig. 4) is widely regarded as a benchmark solution due to its inherent bidirectionality, high power density, and modular structure [7].

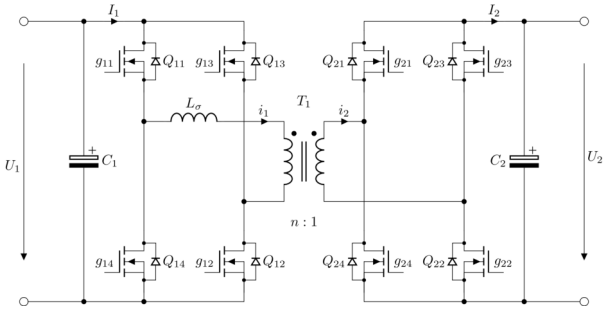


Fig. 4. Dual Active Bridge Topology

A. Dual Active Bridge (DAB)

The DAB converter consists of two active full-bridge stages interconnected by a high-frequency transformer and a series energy-transfer inductor, as illustrated in Fig. 4. Power transfer is regulated by controlling the phase-shift angle between the square-wave voltages generated by the primary and secondary bridges [3]. This phase-shift modulation enables seamless bidirectional power flow and allows simple and robust control implementation.

A key advantage of the DAB topology is its ability to achieve Zero Voltage Switching (ZVS) over a wide operating range without auxiliary circuits, resulting in reduced switching losses and high efficiency. In addition, the symmetrical structure of the DAB supports modular system design and predictable power flow characteristics [7].

Despite these benefits, the DAB exhibits circulating (backflow) power when the input and output voltage ratio deviates from the transformer turns ratio, which increases current stress and degrades efficiency, particularly under light-load conditions [13]. Furthermore, DC magnetic flux bias in the transformer can lead to saturation issues, requiring dedicated compensation or control strategies to ensure reliable operation.

B. Resonant and Hybrid Isolated Topologies

To overcome the limitations of the conventional DAB, recent research has focused on resonant and hybrid isolated topologies. Resonant converters, such as LLC and CLLC (capacitor-inductor-inductor-capacitor) structures, employ resonant tanks to extend soft-switching operation (ZVS/ZCS) and

improve efficiency over wider voltage and load ranges [7], [14]. The LLC converter utilizes a resonant network formed by a series inductance, transformer magnetizing inductance, and a resonant capacitor, enabling efficient energy transfer and wide ZVS operation. In particular, CLLC converters are well suited for bidirectional EV charging applications due to their symmetrical behavior in both power flow directions.

Hybrid isolated topologies combine active bridge structures with resonant elements, such as LCC–DAB or CLLC–DAB configurations, to balance the flexibility of phase-shift control with the efficiency benefits of resonant operation [5]. While these solutions can significantly reduce circulating current and improve peak efficiency, they typically require more complex modulation and control strategies, including variable-frequency operation [1].

Overall, the classic DAB remains a robust choice for applications demanding wide voltage regulation and fast dynamic response, whereas resonant and hybrid converters are increasingly adopted in scenarios where maximum efficiency within a narrower operating range is prioritized.

V. Modulation and Control Strategies

The performance of bidirectional DC–DC converters (BDCs), particularly Dual Active Bridge (DAB) topologies, is strongly influenced by the selected modulation and control strategy [13]. While the converter topology defines the physical structure, the control approach determines efficiency, dynamic response, and the achievable soft-switching operating range [5].

A. Classical Phase-Shift Control

The most widely used modulation technique for DAB converters is Single Phase Shift (SPS), where the transferred power is regulated by adjusting the phase angle between the primary and secondary bridge voltages, as illustrated in Fig. 5 [15]. In conventional implementations, this phase shift is typically controlled using a Proportional–Integral (PI) regulator based on the output voltage error.

Although SPS control is simple and robust, its efficiency deteriorates when the voltage conversion ratio deviates from unity, leading to increased circulating power and higher current stress [13]. Moreover, PI-based control is inherently reactive and exhibits limited performance when dealing with nonlinear operating conditions, such as constant power loads (CPLs), and at light-load operation where the Zero Voltage Switching (ZVS) range is reduced [2], [13].

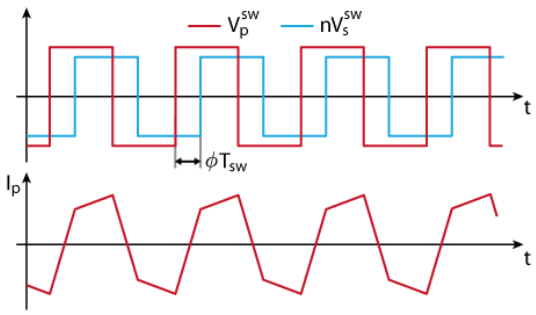


Fig. 5. Phase-shift modulation in DAB converters [16]

B. Advanced Phase-Shift Modulation

To overcome the limitations of SPS, advanced phase-shift modulation techniques such as Extended (EPS), Double (DPS), and Triple Phase Shift (TPS) have been proposed [5]. By introducing additional degrees of freedom, these methods reduce circulating current, extend ZVS operation, and improve efficiency over a wider operating range [13]. However, the increased mathematical complexity and computational requirements, particularly for TPS, limit their practical implementation in real-time control systems with constrained processing resources [13].

C. Model Predictive Control

Model Predictive Control (MPC) represents an advanced control paradigm that enables proactive optimization of converter operation by predicting future system behavior using a mathematical model [2]. MPC is particularly attractive for BDC applications because it can explicitly handle multivariable constraints, such as current limits, voltage boundaries, and battery state-of-charge (SoC), while providing fast dynamic response and improved disturbance rejection [2], [4].

Despite its advantages, MPC suffers from a high computational burden, as the optimization problem must be solved within each switching period. This challenge becomes critical in systems employing Wide Bandgap (WBG) devices, where high switching frequencies significantly reduce the available computation time [8], [17]. Furthermore, MPC performance is highly dependent on model accuracy, and parameter uncertainties may degrade steady-state performance if not properly compensated [8].

VI. Applications and System-Level Perspective

Bidirectional DC–DC converters (BDCs) play an important role in modern power electronic systems by enabling efficient energy exchange between distributed energy resources, storage units, and DC grids. They serve as key components of DC microgrids and energy routing architectures, improving grid stability and operational flexibility [18].

In battery energy storage systems (BESS), BDCs interface energy storage units with the common DC bus and enable efficient power management under varying load conditions [19]. Hybrid energy storage systems (HESS), combining lithium-ion batteries and supercapacitors, benefit from bidirectional converters due to their capability to manage both high energy and high power dynamics [19]. Moreover, control strategies implemented through DC–DC stages can improve system efficiency and extend battery lifetime [20].

In electric vehicle (EV) applications, BDCs connect low-voltage battery systems with high-voltage DC buses used in traction inverters and fast-charging infrastructure [6]. The increasing demand for high-power DC fast charging has accelerated the use of isolated bidirectional topologies, such as the Dual Active Bridge (DAB), together with Wide Bandgap (WBG) semiconductor devices to achieve higher power density and more compact designs [1], [5].

The Vehicle-to-Grid (V2G) concept further expands the role of BDCs by enabling bidirectional power flow between EVs and the grid [4]. With advanced control methods such as Model Predictive Control (MPC), EVs can support grid services including frequency regulation and voltage stabilization, facilitating better integration of renewable energy sources [4].

VII. Research Gaps and Future Directions

Despite significant progress in bidirectional DC–DC converter design, notable gaps remain between theoretical modeling and practical hardware implementation. A major limitation of existing approaches is the reliance on linearized small-signal models, which often fail to accurately describe large-signal dynamics and stability issues caused by constant power loads (CPLs) in DC microgrids [8]. Although Model Predictive Control (MPC) provides an effective framework for handling multivariable constraints, its high computational complexity poses a critical challenge for high-frequency GaN-based converters operating in the MHz range [4].

Another unresolved challenge lies in maintaining Zero Voltage Switching (ZVS) over wide voltage conversion ranges, such as those required in EV fast-charging applications. Conventional DAB converters suffer from increased circulating power and duty-cycle loss when operating far from the nominal voltage ratio, which significantly degrades light-load efficiency and limits overall system performance [13], [14].

From a mid-term perspective, these limitations indicate that further improvements cannot rely solely on advanced control strategies. Instead, there is a growing need for converter topologies that inherently reduce circulating power and improve efficiency over wide operating ranges. In this context, emerging concepts such as partial power converters (PPCs) represent a promising research direction, as they process only a fraction of the total transferred power, enabling higher efficiency and reduced component stress.

Future research is therefore expected to focus on both the development of computationally efficient control methods and the exploration of new converter architectures that combine high efficiency, scalability, and robust operation under dynamic conditions typical for EV and DC microgrid applications.

VIII. Conclusion

In summary, this review confirms that bidirectional DC-DC converters have developed from simple power components into key parts of modern DC grids and electric vehicles. Based on the analyzed literature, it is evident that isolated bridge topologies, particularly the Dual Active Bridge (DAB), offer the most reliable balance of safety and performance for high-power integration.

While the adoption of Wide Bandgap materials like GaN and SiC is drastically increasing switching frequencies and reducing system size, the real breakthrough now lies in the control layer. Analysis indicates that the future of the field depends on moving beyond traditional linear controllers toward proactive, non-linear strategies like Model Predictive Control. Achieving high efficiency over wide voltage ranges requires close cooperation where high-speed hardware is managed by equally agile, software-defined control algorithms.

At the same time, emerging converter concepts such as partial power processing architectures indicate a promising direction for further improving efficiency and reducing system stress. This suggests that future developments will likely combine advanced control strategies with fundamentally new converter structures to meet the demands of next-generation energy systems.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under the project APVV-23-0521 and VEGA 1/0584/24.

References

[1] "Gan-based dc-dc converters for ev fast charging: A review of wide bandgap devices technology," *Results in Engineering*, vol. 28, p. 107548, 2025.

[2] A. Surendar, "Model predictive control of bidirectional converters in grid-interactive battery systems," *Transactions on Power Electronics and Renewable Energy Systems*, vol. 1, no. 1, pp. 13–20, Jan–Mar 2025.

[3] "Isolated bidirectional dc-dc converter: A topological review," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, p. 100594, 2024.

[4] N. Patcharaprakiti, E. Siriboonpanich, and J. Patcharaprakiti, "A bidirectional V2G charging point operator with model predictive control for energy trading platform," in *EVS38 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, Göteborg, Sweden, June 15–18 2025.

[5] S. Qazi, P. Venugopal, A. J. Watson, P. Wheeler, and T. B. Soeiro, "A parallel input and versatile output dual active bridge converter," *IEEE Transactions on Power Electronics*, vol. 40, no. 4, pp. 6203–6215, April 2025.

[6] A. Ahmad, M. M. Reza, A. R. Beig, J. Y. Alsawalhi, and K. A. Jaafari, "High voltage gain switched-z-source bidirectional dc-dc converter," *IEEE Access*, vol. 10, pp. 53 560–53 577, 2022.

[7] A. A. Chamyousefali, "Analysis of dual active bridge converter," Master's Thesis, Politecnico di Milano, Milan, Italy, 2024, academic Year 2024-2025.

[8] Q. Xu, N. Vafamand, L. Chen, T. Dragičević, L. Xie, and F. Blaabjerg, "Review on advanced control technologies for bidirectional dc/dc converters in dc microgrids," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 2, pp. 1205–1221, April 2021.

[9] S. Dharmasena, T. O. Olowu, and A. I. Sarwat, "Bidirectional ac/dc converter topologies: A review," in *2019 SoutheastCon*, April 2019, pp. 1–5.

[10] "A novel multi-port high-gain bidirectional dc–dc converter for energy storage system integration with dc microgrids," *Journal of Energy Storage*, vol. 87, p. 111431, 2024.

[11] Y. Zhang, Y. Gao, L. Zhou, and M. Sumner, "A switched-capacitor bidirectional dc–dc converter with wide voltage gain range for electric vehicles with hybrid energy sources," *IEEE Transactions on Power Electronics*, vol. 33, no. 11, pp. 9459–9469, Nov 2018.

[12] G. Madhumitha, R. Uppara, and S. Priya, *Comparative Analysis of Bidirectional Buck/Boost DC-DC Converter and Interleaved Bidirectional Buck/Boost DC-DC Converter for EV applications*. River Publishers, 2024, ch. 8.

[13] Z. Qin, Y. Shen, P. C. Loh, H. Wang, and F. Blaabjerg, "A dual active bridge converter with an extended high-efficiency range by dc blocking capacitor voltage control," *IEEE Transactions on Power Electronics*, vol. 33, no. 7, pp. 5949–5966, July 2018.

[14] Y. Wang, Y. Shi, and K. Xu, "A new zero-voltage zero-current switching converter with minimum duty cycle loss," *Electronics*, vol. 13, no. 3, 2024.

[15] J. Everts, J. Van den Keybus, F. Krismer, J. Driesen, and J. W. Kolar, "Switching control strategy for full zvs soft-switching operation of a dual active bridge ac/dc converter," in *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, Feb 2012, pp. 1048–1055.

[16] J. Orsinger, "Dab converter control using phase-shift modulation," 2022, posted March 3, 2022; updated May 7, 2025. [Online]. Available: https://imperix.com/doc/implementation/dab-converter-control

[17] S. R. Raman, K. W. E. Cheng, and Y. Ye, "Multi-input switched-capacitor multilevel inverter for high-frequency ac power distribution," *IEEE Transactions on Power Electronics*, vol. 33, no. 7, pp. 5937–5948, July 2018.

[18] H. Jahanghiri, S. Rahimi, A. Shaker, and A. Ajami, "A high conversion non-isolated bidirectional dc-dc converter with low stress for microgrid applications," in *2019 10th International Power Electronics, Drive Systems and Technologies Conference (PEDSTC)*, Feb 2019, pp. 775–780.

[19] "Sliding-mode and lyapunov function-based control for battery/supercapacitor hybrid energy storage system used in electric vehicles," *Energy*, vol. 122, pp. 601–612, 2017.

[20] "Active dc to dc converter based battery charge balancing systems from renewable energy by using electric vehicle," *Energy Reports*, vol. 14, pp. 1114–1136, 2025.

Decomposing the Programmer Attribution Problem in Educational Source Code

[1]*Marek HORVÁTH (3rd year),*
Supervisor: [2]*Emília PIETRIKOVÁ*

[1,2]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.horvath@tuke.sk, [2]emilia.pietrikova@tuke.sk

***Abstract*—Programmer attribution based on source code is commonly formulated as a single classification task, despite being influenced by multiple interacting methodological factors. In educational settings, this simplification is particularly problematic due to task constraints, shared solution patterns, and heterogeneous levels of programming experience. This paper adopts a decompositional view of programmer attribution and examines selected methodological aspects that are currently the most developed within the ongoing research. The work summarizes recent efforts related to feature space structuring, representation and preprocessing choices, and the analysis of similarity-based evidence in educational source code. Particular attention is given to how multiple similarity views can support the interpretation of pairwise relations and anomalous submissions. We reflect the current stage of the research and outline how these partial analyses contribute to the preparation of subsequent integrated experiments.**

***Keywords*—Authorship attribution, Authorship Verification, Behavioral biometrics, Code Forensics, Programmer identification, Programming Education, source code stylometry, Source code similarity, Version control analysis**

I. Introduction

This work addresses programmer attribution in educational source code and focuses on the formulation and structuring of its underlying methodological components. Educational programming courses produce large collections of student submissions that share assignments, constraints, and common solution patterns, which complicates the use of stylistic and behavioral information for attribution oriented analysis.

Previous work has applied source code analysis to several related tasks in educational settings, including the comparison of code similarity tools, plagiarism detection, analysis of code reuse, and the study of stylistic consistency in student programs [1]. Further studies have addressed automated assessment systems and learning analytics pipelines in large programming courses, where scalable analysis of source code is required [2]. More recent work has examined programmer identification using machine learning methods and the presence of AI generated code in student assignments [3].

Most existing approaches treat programmer attribution as a single classification task. Such formulations obscure the fact that attribution relies on a set of distinct methodological components, including feature design, representation, preprocessing, and similarity analysis. This paper adopts a decompositional view of programmer attribution and reports on the current stage of work focused on developing and refining selected components rather than integrating them into a unified attribution model. The presented analyses serve as a preparatory step toward later integrated experimental evaluation.

A. Terminology

Core terms are sometimes used with different scopes or implicit assumptions depending on the research context. To avoid ambiguity, this section introduces working definitions of key concepts as they are used in this paper, with a focus on educational source code.

Authorship attribution. Authorship attribution refers to the task of determining who wrote a given piece of text or source code. In the context of software forensics, it represents the process of analyzing programming artifacts to identify their likely author [4]. The goal is to associate a code fragment with one programmer among a known set of candidates, typically using features extracted from the code itself.

Stylometry. Stylometry is the quantitative study of individual writing style. In programming, it examines how personal habits in code layout, naming conventions, control structures, and syntactic organization manifest as measurable stylistic patterns [5]. Stylometric features are commonly grouped as lexical, syntactic, structural, or semantic, and may be extracted through tokenization or analysis of abstract syntax trees (ASTs), which capture the hierarchical structure of source code.

Behavioral biometrics. Behavioral biometrics focus on identifying individuals through distinctive behavioral patterns, such as typing dynamics, edit sequences, or commit activity. In software engineering, such data capture temporal and interactive aspects of programming that complement static code stylometry [6].

II. Related Work

This section summarizes selected observations derived from the preparation of a recent review article on programmer attribution in source code. During this review, a broad literature search was conducted to identify studies addressing authorship attribution, verification, and programmer identification using stylometric and behavioral analysis. Rather than providing a full survey, this section highlights recurring experimental choices and trends that motivated the decomposition adopted in this work.

The complete set of keywords used in the search process is presented in Table I. Across all screening stages, a total of 135 records were examined in detail, of which 47 were

included and 88 were excluded based on the defined selection criteria. The screening and extraction process was supported by a custom parsing script[1].

TABLE I: Keyword Groups Used in the Search Process

Group	Focus	Keywords used in search
A	Authorship task	"authorship attribution" "authorship verification" "programmer identification" "developer identification"
B	Source code object	"source code" "programming assignments" "software repositories" "GitHub"
C	Analytical approach	"stylometry" "behavioral biometrics" "commit behavior" "interaction logs" "code analysis"

Datasets and programming languages play a central role in the reproducibility and generality of programmer attribution research. Figure 1 groups the included studies according to their primary data source. The *Google Code Jam* dataset is the most frequently used benchmark, appearing in 25 papers, mainly due to its size, public availability, and standardized programming tasks. GitHub repositories form the second largest category with 15 studies and are typically used to capture development styles from open source projects [7]. The remaining 14 studies rely on synthetic or custom datasets, reflecting the need to adapt data to specific experimental settings or research goals.

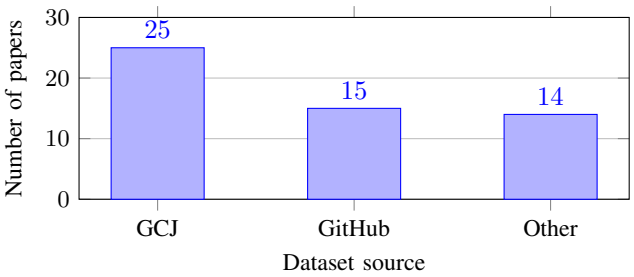


Fig. 1: Dataset source categories for the included studies.

Programming language coverage further illustrates common practices in the reviewed literature. Among the analyzed studies, 28 do not specify a concrete programming language, which is often associated with language independent features or multi language datasets. C++ [8], [9], [10], [11] and Java [12], [4], [13], [14] each appear in four papers, while Python is considered in two studies [15], [16]. Nine studies explicitly employ mixed language datasets. An overview of programming languages and code representations observed in the reviewed studies is provided in Table II.

Code representation choices determine how stylistic and structural information is extracted from source code. Abstract syntax trees are the most commonly used representation, appearing in six studies [17], [15], [18], [19], [4], [20]. Embedding based representations are employed in three studies [8], [21], [20], while token based approaches appear in two papers [18], [12]. Concrete syntax trees [15] and program dependence graphs [22] are each used in a single study. The frequent use of abstract syntax trees reflects a preference for structured representations that preserve syntactic hierarchy while remaining applicable across different programming languages.

[1] https://github.com/marek-horvath/survey-extracting-script

TABLE II: Occurrence of programming languages and code representations in the included studies.

Category	Subcategory	Number of papers
Programming languages	Unspecified	28
	Mixed	9
	C++	4
	Java	4
	Python	2
Code representations	AST	6
	Embeddings	3
	Tokens	2
	CST	1
	PDG	1

III. Behavioral Components from Version Control

Behavioral components are derived from version control activity in educational programming assignments and are treated as a separate dimension of the authorship attribution problem. Behavioral data are extracted from Git repositories associated with student coursework and analyzed independently of static source code features. The considered signals capture temporal and activity related aspects of the programming process, including commit frequency, timing with respect to assignment deadlines, magnitude of code changes, and basic characteristics of commit messages. To support this analysis, dedicated tooling was developed to automate data extraction, aggregation, and feature computation across educational repositories.

The observed behavioral patterns show substantial variation among students, even within the same assignment and programming language. Some students exhibit gradual development with frequent small commits distributed across the working period, while others concentrate their activity close to deadlines with larger changes. These differences indicate that behavioral components capture process related characteristics that are not directly observable from final code snapshots alone.

Within the decomposed research framework, behavioral components are not used in isolation for attribution. Instead, they serve as interpretable building blocks intended for later integration with static representations, helping to separate process related effects from structural properties of source code.

IV. Graph Exploration of Similarity and Outliers

Pairwise similarity relations among student submissions are analyzed using a graph representation to support the decomposition of authorship related evidence. The focus is not on direct author attribution, but on examining how different similarity signals interact and how they structure the solution space in educational programming tasks.

The analysis is based on several decomposed similarity components:

- token similarity computed using Jaccard similarity on raw and normalized token representations to separate lexical overlap from identifier renaming effects,
- AST similarity capturing structural correspondence between programs independently of formatting,
- composite similarity thresholds requiring agreement across multiple similarity views before establishing pairwise relations,
- outlier detection based on structural code metrics to identify atypical submissions that deviate from dominant solution patterns.

Student submissions are represented as nodes in a similarity graph, with edges denoting pairs exceeding predefined similarity thresholds. This representation enables the identification of dense clusters, isolated high similarity pairs, and structurally atypical submissions. Such patterns provide complementary evidence for interpreting similarity relations in large cohorts without relying on a single similarity score.

V. Integrated Experimental Setup

The next stage of the research focuses on applying the developed components to real educational datasets and using them for programmer attribution tasks. Static stylometric features, similarity based measures, and behavioral data extracted from version control systems will be combined within a single experimental setup and evaluated jointly. The experiments will use standard machine learning classifiers, primarily Random Forest and XGBoost, which are suitable for heterogeneous feature sets commonly encountered in educational source code data. In addition to attribution, the same setup will be used to identify outlier submissions that deviate from typical solution patterns or show inconsistencies between static and behavioral characteristics. The emphasis of this phase is on consistent preprocessing, stable data splits, and evaluation across assignments and courses.

VI. Conclusion

This paper described the current stage of research that approaches programmer attribution in educational source code as a decomposable problem. Based on observations from a review of existing work, the focus was placed on preparing and refining selected components rather than constructing a single end to end attribution model. The next step is to apply these components jointly to collected educational data and to evaluate their behavior in practical attribution and outlier detection scenarios.

References

[1] M. Horváth and E. Pietriková, "An experimental comparison of three code similarity tools on over 1,000 student projects," in *2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2024, pp. 000 423–000 428.

[2] M. Horváth, E. Pietriková, and F. Gurbál', "Personalized learning analytics through static code analysis in computer science education," *Acta Informatica Pragensia*, vol. 2026, no. 1, pp. 54–71, 2026.

[3] M. Horváth, M. Čupalka, and E. Pietriková, "Detecting programmer identity from source code with machine learning methods," in *2025 International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 2025, pp. 236–241.

[4] M. F. Tennyson, "A Replicated Comparative Study of Source Code Authorship Attribution," in *2013 3rd International Workshop on Replication in Empirical Software Engineering Research*, Oct. 2013, pp. 76–83. [Online]. Available: https://ieeexplore.ieee.org/document/6664734/

[5] A. Gurioli, M. Gabbrielli, and S. Zacchiroli, "Stylometry for real-world expert coders: a zero-shot approach," *PeerJ Computer Science*, vol. 10, 2024.

[6] A. Caliskan-Islam, R. Harang, A. Liu, A. Narayanan, C. Voss, F. Yamaguchi, and R. Greenstadt, "De-anonymizing programmers via code stylometry," in *Proc. USENIX Secur. Symp.* USENIX Association, 2015, pp. 255–270.

[7] E. Chatzicharalampous, G. Frantzeskou, and E. Stamatatos, "Author Identification in Imbalanced Sets of Source Code Samples," in *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, vol. 1, Nov. 2012, pp. 790–797. [Online]. Available: https://ieeexplore.ieee.org/document/6495124/

[8] M. Abuhamad, J.-s. Rhim, T. AbuHmed, S. Ullah, S. Kang, and D. Nyang, "Code authorship identification using convolutional neural networks," *Future Generation Computer Systems*, vol. 95, pp. 104–115, 2019.

[9] S. Alrabaee, M. Debbabi, and L. Wang, "CPA: Accurate Cross-Platform Binary Authorship Characterization Using LDA," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3051–3066, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9032128/

[10] P. Bayrami and J. E. Rice, "Code Authorship Attribution using content-based and non-content-based features," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Sep. 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9569061/

[11] S. HajiHosseinKhani, B. Joshi, and A. Habibi Lashkari, "AuthAttLyzer-V2: Unveiling Code Authorship Attribution Using Enhanced Ensemble Learning Models and Generating Benchmark Dataset," in *Commun. Comput. Info. Sci.*, vol. 2253 CCIS. Springer Science and Business Media Deutschland GmbH, 2025, pp. 148–165.

[12] B. Al-Ahmad, N. Al-Madi, A. Alzaqebah, R. S. Alkhawaldeh, K. Aldebei, M. F. Kabir, I. Altaharwa, M. Abu-Faraj, and I. Aljarah, "Meta-Heuristic Guided Feature Optimization for Enhanced Authorship Attribution in Java Source Code," *IEEE Access*, vol. 11, pp. 141 657–141 673, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10352131

[13] W. Wang, G. Meng, H. Wang, K. Chen, W. Ge, and X. Li, "A3Ident: A Two-phased Approach to Identify the Leading Authors of Android Apps," in *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, Sep. 2020, pp. 617–628. [Online]. Available: https://ieeexplore.ieee.org/document/9240707/

[14] R. Yokomori and K. Inoue, "An Empirical Analysis of Code Clone Authorship in Apache Projects," in *2023 IEEE 17th International Workshop on Software Clones (IWSC)*, Oct. 2023, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/10473617/

[15] S. Balla, M. Gabbrielli, and S. Zacchiroli, "Code stylometry vs formatting and minification," *PeerJ Computer Science*, vol. 10, 2024.

[16] P. Hozhabrierdi, D. Fuentes Hitos, and C. K. Mohan, "Python source code de-anonymization using nested bigrams," in *IEEE Int. Conf. Data Mining Workshops, ICDMW*, vol. 2018-November. IEEE Computer Society, 2018, pp. 23–28.

[17] B. Alsulami, E. Dauber, R. Harang, S. Mancoridis, and R. Greenstadt, "Source code authorship attribution using long short-term memory based networks," in *Lect. Notes Comput. Sci.*, vol. 10492 LNCS. Springer Verlag, 2017, pp. 65–82.

[18] B. R. Cernansky and I. J. Petrik, "Comparing Machine Learning and Deep Learning Models With Attention Mechanisms for Multiple Source Code Authorship Attribution," in *Comput. Syst. Technol. - Int. Conf., CompSysTech - Proc.* Institute of Electrical and Electronics Engineers Inc., 2025.

[19] J. Petrik and D. Chuda, "The effect of time drift in source code authorship attribution: Time drifting in source code - stylochronometry," in *International Conference on Computer Systems and Technologies '21*. Ruse Bulgaria: ACM, Jun. 2021, pp. 87–92. [Online]. Available: https://dl.acm.org/doi/10.1145/3472410.3472445

[20] S. Zafar, M. U. Sarwar, S. Salem, and M. Z. Malik, "Language and Obfuscation Oblivious Source Code Authorship Attribution," *IEEE Access*, vol. 8, pp. 197 581–197 596, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9245552/

[21] R. White and N. Sprague, "Deep Metric Learning for Code Authorship Attribution and Verification," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2021, pp. 1089–1093. [Online]. Available: https://ieeexplore.ieee.org/document/9680230

[22] F. Ullah, J. Wang, S. Jabbar, F. Al-Turjman, and M. Alazab, "Source Code Authorship Attribution Using Hybrid Approach of Program Dependence Graph and Deep Learning Model," *IEEE Access*, vol. 7, pp. 141 987–141 999, 2019.

Sensorless PMSM Control via High-Frequency Signal Injection: Principles and Methods

[1]*Peter ŠOFRANKO (1st year),*
Supervisor: [2]*Karol KYSLAN*

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Košice, Slovak Republic

[1]peter.sofranko@tuke.sk, [2]karol.kyslan@tuke.sk

***Abstract*—This paper provides an overview of sensorless control techniques for permanent magnet synchronous motors (PMSMs), with a focus on high-frequency (HF) signal injection methods for low- and zero-speed operation. Sensorless techniques can be divided into model-based methods, used at medium and high speeds, and saliency-based methods, which are suitable for low speeds. At low speeds, the back electromotive force (EMF) is too small for reliable position estimation, which limits the performance of conventional approaches. For this reason, HF signal injection methods are a promising solution. The paper presents and compares several injection techniques, including rotating and pulsating sinusoidal, square-wave, and pulse-wave signals. Special attention is given to the HF pulse-wave injection method as a promising approach for robust and accurate rotor position estimation at low and zero speeds.**

***Keywords*—High-frequency signal injection, permanent magnet synchronous motor, position estimation, sensorless control.**

I. Introduction

In industry, synchronous motors are used in devices and applications with different speed and power ranges. Permanent magnet synchronous motors (PMSMs) are usually applied in small and medium-power applications, especially in servo drives, robotics, and motion-control devices [1]. The type of control depends on the specific application. Field-oriented control (FOC) and direct torque control (DTC) are the most common methods for PMSM control in industrial applications [2]. PMSMs can operate over the entire speed range, from the lowest to the maximum speeds, and can use different control methods, in general, sensored and sensorless control methods. In the medium to high speed range, sensorless control is typically applied because the control methods are well established and widely used. At low speeds, these methods are limited, so alternative control structures and techniques are being developed [3]. The high-frequency (HF) voltage signal injection method is particularly effective at low speeds. The injected HF signal is in most cases superimposed onto fundamental control signals and this injection is reflected in specific current response [4]. From that current response, the rotor speed and position can easily be obtained. The use of sensorless control in the low-speed range is standard in industrial applications, electric vehicle applications, household appliances, as well as in fan and pump systems [1], [3].

II. Sensorless control

The knowledge of the rotor angular velocity and position is essential in control system of PMSM motor drives. This information is typically obtained using various types of speed and position sensors. Many of these sensors are unreliable because they are sensitive to electromagnetic interference. They are often expensive, require additional installation space near the motor, and need cable connections [2]. At present, the field of sensorless control is developing rapidly. The concept of sensorless control is based on measuring motor electrical quantities, from which the rotor position and speed can be estimated. Most commonly, the motor voltages and currents are measured [5].

The typical classification of sensorless control methods includes [3], [6], [7], [8], [9]:

1) Model-Based Methods (Electromotive Force or Flux Estimation, Kalman Filters, Sliding Mode Observers).
2) Artificial Intelligence-Based Methods (Fuzzy Logic, Neural Networks, Fuzzy Neural Networks).
3) Saliency-Based Methods (HF Signal Injection).

III. Saliency-Based sensorless control methods

At low motor speeds, the electromotive force (EMF) is low and insufficient for reliable control. In this range, the EMF value is nearly equal to the noise level, making it difficult to distinguish [10]. For this low-speed operating range (5–10 % of the motor's rated speed) a HF signal injection method is used. This method is based on the magnetic asymmetry of the motor – the motor inductance along the d-axis L_d differs at a given instant from the inductance along the q-axis L_q. This asymmetry is advantageous for operation at low to zero speeds. In most cases, a HF voltage signal is injected and the resulting current response is observed. Saliency-based (signal injection-based) methods apply rotating sinusoidal, pulsating sinusoidal, pulsating square-wave, and pulsating pulse-wave signals for HF injection [1], [3], [4]. One of the most recently described methods for the low-speed range is the pseudo-random signal injection method, which eliminates all discrete harmonics in the noise spectrum, resulting in effective noise attenuation [1], [11], [12]. These signals are injected into a modified PMSM model to perform HF signal injection.

The PMSM model in the stationary reference frame is a simplified representation obtained by transforming the three-phase a, b, c model into a two-phase $\alpha\beta$ model. The reference frame $\alpha\beta$ is fixed relative to the stator and is transformed into the synchronous reference frame dq, which is fixed to the rotor. The angle θ_e denotes the angle between these reference frames. The estimated synchronous reference frame $\hat{d}\hat{q}$ is

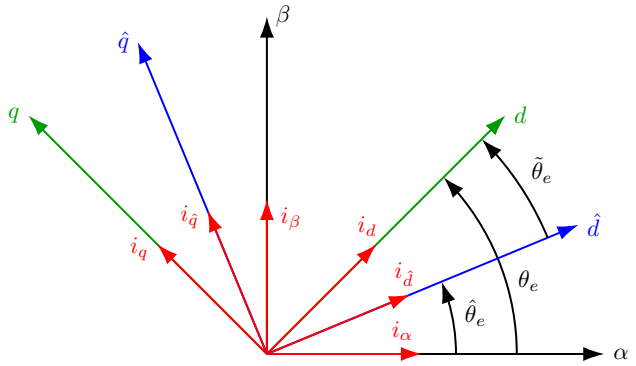


Fig. 1. Reference frames $\alpha\beta$, dq, and $\hat{d}\hat{q}$; angles θ_e, $\hat{\theta}_e$, and $\tilde{\theta}_e$ between these frames; currents $i_{\alpha\beta}$, i_{dq}, and $i_{\hat{d}\hat{q}}$ expressed in their respective reference frames.

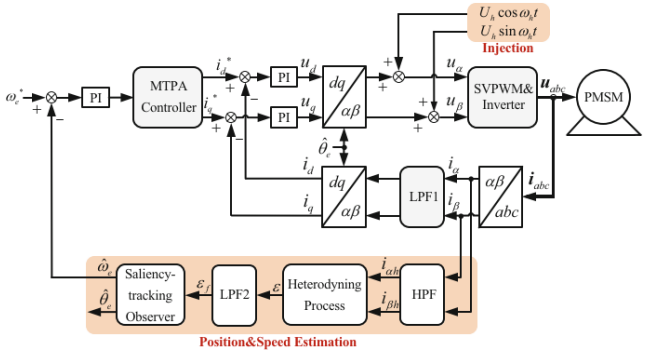


Fig. 2. Block diagram of HF rotating sinusoidal signal injection into the stationary reference frame $\alpha\beta$ [1].

derived from the synchronous reference frame dq. In this case, the angle $\tilde{\theta}_e$ denotes the angle between the dq and $\hat{d}\hat{q}$ reference frames, while the angle $\hat{\theta}_e$ denotes the angle between the stationary reference frame $\alpha\beta$ and the estimated synchronous reference frame $\hat{d}\hat{q}$. The reference frames are shown in Figure 1.

In general, HF signals are injected into the stationary reference frame $\alpha\beta$ or the estimated synchronous reference frame $\hat{d}\hat{q}$ [1], [4]. The objective is the measurement of the current response during HF signal injection.

From the general PMSM model, a modified PMSM model for performing HF signal injection is derived [1], [4]. Based on the modified PMSM model, the change in currents i_α and i_β under HF signal injection in the stationary reference frame $\alpha\beta$ or in the estimated synchronous reference frame $\hat{d}\hat{q}$ can be expressed as shown in (1) and (2). Equations (1) and (2) contain information about the rotor position, and therefore, by observing the current change, it is also possible to determine the change in rotor position.

$$\frac{d}{dt}\begin{bmatrix} i_{\alpha h} \\ i_{\beta h} \end{bmatrix} = \frac{1}{\Sigma L^2 - \Delta L^2}\left[\Sigma L \begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} - \Delta L \begin{bmatrix} \cos 2\theta_e + \sin 2\theta_e \\ \sin 2\theta_e - \cos 2\theta_e \end{bmatrix} \begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} \right], \tag{1}$$

$$\frac{d}{dt}\begin{bmatrix} i_{\alpha h} \\ i_{\beta h} \end{bmatrix} = \frac{1}{\Sigma L^2 - \Delta L^2}\left[\Sigma L \begin{bmatrix} \cos\hat{\theta}_e & -\sin\hat{\theta}_e \\ \sin\hat{\theta}_e & \cos\hat{\theta}_e \end{bmatrix} \begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} - \Delta L \begin{bmatrix} \cos(2\theta_e - \hat{\theta}_e) & \sin(2\theta_e - \hat{\theta}_e) \\ \sin(2\theta_e - \hat{\theta}_e) & -\cos(2\theta_e - \hat{\theta}_e) \end{bmatrix} \begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} \right], \tag{2}$$

where $u_{\alpha h}$, $u_{\beta h}$, $i_{\alpha h}$, $i_{\beta h}$ are HF voltages and currents in the stationary reference frame, $\hat{\theta}_e$ is the estimated position via the saliency-tracking observer, $u_{\hat{d}h}$ and $u_{\hat{q}h}$ are HF voltages in the estimated reference frame. Inductances ΣL and ΔL represent the average and differential inductances and are defined as:

$$\Sigma L = \frac{L_d + L_q}{2}, \qquad \Delta L = \frac{L_d - L_q}{2}, \tag{3}$$

where L_d and L_q are the inductances along the d- and q-axes.

In both cases, after injecting HF signals and measuring the current response, these responses are different. Depending on the reference frame used, the responses must be demodulated and the resulting position-dependent signal is used as the input to a position observer to estimate rotor speed and position. Most commonly, a phase-locked loop (PLL) is used for this purpose [3], [4], [13].

A. High Frequency Rotating Sinusoidal Signal Injection

This method was described in [14] in 1995, and the block diagram of the injection is shown in Figure 2. A rotating voltage space vector is injected in $\alpha\beta$-axes (4):

$$\begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} = U_h \begin{bmatrix} \cos(\omega_h t) \\ \sin(\omega_h t) \end{bmatrix}, \tag{4}$$

where U_h and ω_h are the amplitude and angular speed of the injected HF voltage signal [1].

The current response is composed of two main components: a positive sequence i_p and a negative sequence i_n. These current responses are obtained using band-pass filters. The positive-sequence component of the response is shifted to the HF domain and filtered by a low-pass filter. The negative-sequence component contains the rotor position information and can be estimated from it [4]. The disadvantages of this method are low injection frequency, distortion of the injected signals and phase modulation. The main advantage of this method is that no initial position information is required [3].

B. High Frequency Pulsating Sinusoidal Signal Injection

The previously used injection method and signal demodulation process are relatively complex to implement, and the use of filters adversely affects dynamic performance. Another method injects a HF pulsating sinusoidal voltage signal into the stationary reference frame $\alpha\beta$, where the signal is applied along the α-axis (5). The HF current response is amplitude-modulated, which simplifies the signal demodulation process and results in improved dynamic performance compared to rotating sinusoidal signal injection [4].

$$\begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} = U_h \begin{bmatrix} \cos(\omega_h t) \\ 0 \end{bmatrix} \tag{5}$$

If a HF pulsating sinusoidal voltage signal is injected into the β-axis, the injection can be expressed by (6). Similar to the α-axis injection, the resulting current response is amplitude-modulated by the rotor position and the demodulation process is identical [4].

$$\begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} = U_h \begin{bmatrix} 0 \\ \cos(\omega_h t) \end{bmatrix} \tag{6}$$

HF pulsating sinusoidal voltage signal injection into the estimated $\hat{d}$-axis was described in [15] in 1998 and the

characteristics of the injected signals are defined in (7). In [1], [3], a sine function is used as the injected signal, while in [15] a cosine function is applied. The structure and model of injection into the estimated $\hat{d}$-axis are described in [1].

$$\begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} = U_h \begin{bmatrix} \cos(\omega_h t) \\ 0 \end{bmatrix} \tag{7}$$

In the demodulation process, the stator currents in the stationary reference frame $\alpha\beta$ are transformed into the estimated synchronous reference frame $\hat{d}\hat{q}$, which also includes the estimated rotor position $\hat{\theta}_e$. The amplitude of the q-axis HF current response is related to the position error. The position error is denoted as $\Delta\theta$ [4] or $\tilde{\theta}_e$ [1] and represents the error between the synchronous reference frame dq and the estimated synchronous reference frame $\hat{d}\hat{q}$, as shown in Figure 1. This position error is expressed as $\tilde{\theta}_e = \theta_e - \hat{\theta}_e$, where θ_e represents the angle between stationary reference frame $\alpha\beta$ and synchronous reference frame dq, $\hat{\theta}_e$ represents the angle between stationary reference frame $\alpha\beta$ and estimated synchronous reference frame $\hat{d}\hat{q}$.

C. High Frequency Pulsating Square-Wave Injection

The structure of pulsating square-wave injection is similar to the pulsating sinusoidal signal injection scheme. In this case, the injected frequency can be much higher than the rotating injection or pulsating sinusoidal injection schemes. The injection of a HF square-wave signal into the stationary reference frame $\alpha\beta$ is described in [4], [16], where the signal is applied along the α or β-axis. An injection into the α-axis (8) can be expressed as:

$$\begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} = \begin{bmatrix} U_h(-1)^n \\ 0 \end{bmatrix}, \tag{8}$$

where U_h is the amplitude and n is an integer and the number of injected voltages.

A similar definition is used in the estimated synchronous reference frame $\hat{d}\hat{q}$ (9), where a pulsating square-wave signal is injected into the estimated $\hat{d}$-axis [1], [17], [18].

$$\begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} = \begin{bmatrix} U_h(-1)^k \\ 0 \end{bmatrix}, \tag{9}$$

where U_h and k indicate the amplitude and sequence of the injected voltage. Square-wave injection in the estimated synchronous reference frame $\hat{d}\hat{q}$ is illustrated in Figure 3. The advantages of the pulsating square-wave signal injection method compared to pulsating sinusoidal signal injection include a higher injection frequency and smaller HF current responses, resulting in reduced additional noise. The disadvantage of this method is that the square-wave signals cause high losses [1].

D. High Frequency Pulsating Pulse-Wave Injection

The HF pulse-wave signal injection method was described for the first time in 2020 in [1]. Although the pulse-wave method is very similar to the square-wave injection method, experiments indicate that it offers several advantages. General block diagram of pulse-wave signal injection is shown in Figure 4.

The injection method in the estimated synchronous reference frame $\hat{d}\hat{q}$ uses a pulse-wave signal, which consists of positive (10) and negative (11) injection periods. The subscripts p and n indicate the polarity of the pulse-wave signal. This positive and negative injection periods are applied to the estimated $\hat{d}$-axis, while zero is applied to the estimated $\hat{q}$-axis.

$$\begin{bmatrix} u_{\hat{d}h-p} \\ u_{\hat{q}h-p} \end{bmatrix} = \begin{bmatrix} U_h \\ 0 \end{bmatrix}, \tag{10}$$

$$\begin{bmatrix} u_{\hat{d}h-n} \\ u_{\hat{q}h-n} \end{bmatrix} = \begin{bmatrix} -U_h \\ 0 \end{bmatrix}, \tag{11}$$

where U_h is the amplitude of injected voltages.

The HF current responses are derived from equation (2), and the HF current derivative di/dt is approximated by $\Delta i/\Delta t$. Also, the HF voltages $u_{\hat{d}h}$, $u_{\hat{q}h}$ in the estimated reference frame are replaced by the expressions in (10) and (11). For this type of injected signal, the HF current responses are described in (12) and (13).

$$\begin{bmatrix} \Delta i_{\alpha h-p} \\ \Delta i_{\beta h-p} \end{bmatrix} = \frac{U_h \Delta t}{L_d L_q} \begin{bmatrix} L_d \sin(\theta_e)\sin(\tilde{\theta}_e) \\ -L_d \cos(\theta_e)\sin(\tilde{\theta}_e) \\ +L_q \cos(\theta_e)\cos(\tilde{\theta}_e) \\ +L_q \sin(\theta_e)\cos(\tilde{\theta}_e) \end{bmatrix}, \tag{12}$$

$$\begin{bmatrix} \Delta i_{\alpha h-n} \\ \Delta i_{\beta h-n} \end{bmatrix} = \frac{-U_h \Delta t}{L_d L_q} \begin{bmatrix} L_d \sin(\theta_e)\sin(\tilde{\theta}_e) \\ -L_d \cos(\theta_e)\sin(\tilde{\theta}_e) \\ +L_q \cos(\theta_e)\cos(\tilde{\theta}_e) \\ +L_q \sin(\theta_e)\cos(\tilde{\theta}_e) \end{bmatrix}, \tag{13}$$

where Δt denotes sampling period, θ_e denotes the rotor position and $\tilde{\theta}_e$ denotes the position error.

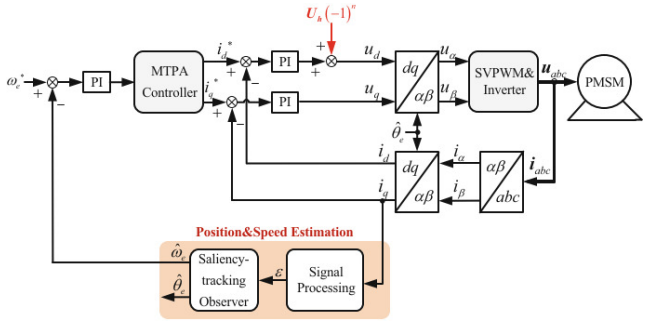


Fig. 3. Block diagram of HF square-wave signal injection. After injecting the HF signal into the estimated $\hat{d}$-axis, the current i_q contains rotor position information and is measured. The Signal Processing block calculates the estimation error ϵ, which is fed into the PI controller in the Saliency-tracking Observer; the output of the observer is the estimated rotor position $\hat{\theta}_e$ [1].

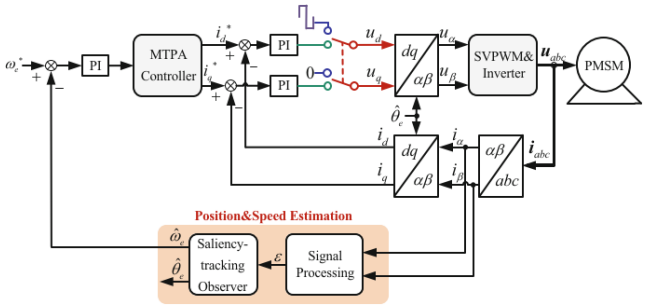


Fig. 4. Block diagram of the HF pulse-wave signal injection method. The current response is measured, and the current samples i_α, i_β are processed in the Signal Processing block. The estimation of angular speed $\hat{\omega}_e$ and rotor position $\hat{\theta}_e$ in the Saliency-tracking Observer block is identical to that shown in Figure 3 [1].

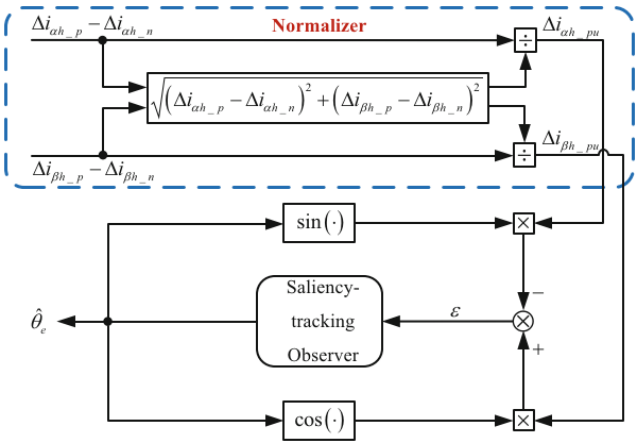


Fig. 5. Process of rotor position and angular speed estimation. After normalization of the current samples, the estimation error ϵ is processed in the Saliency-tracking Observer block. This observer uses a simplified PLL, which employs a PI controller. The input signal (control error) for the PI controller is the estimation error ϵ, which, in the low-speed range, is identical to the position error $\tilde{\theta}_e$. The output of the controller is the estimated angular speed $\hat{\omega}_e$, and after integration, the estimated rotor position $\hat{\theta}_e$ is obtained [1].

The current sampling is defined by the injection sequence. Each control period is divided into a FOC period and an injection period, which are always separated. When the FOC period is active, no HF signal is injected. During the injection period, the current response sample is an induced HF current. The current sample is measured at the beginning of each switching period. Every sample of the current response from both the positive and negative injection periods is used to estimate the rotor speed and position. However, the obtained samples of the current responses are affected by the effects of inductance variation and fluctuations in the amplitude of the injected signal. In this case, normalization is applied. Then, the estimation error ϵ is calculated and used as the input to the saliency-tracking observer (PLL) [1]. Finally, the output of the entire process is the estimated angular velocity $\hat{\omega}_e$ and rotor position $\hat{\theta}_e$ [1], [13]. The entire process of normalization and the calculation of the estimated angular velocity $\hat{\omega}_e$ and rotor position $\hat{\theta}_e$ is illustrated in Figure 5.

IV. Conclusion

This paper describes the main methods for sensorless control of PMSMs. Despite limitations, sensorless control offers well-developed techniques for rotor speed and position estimation, particularly in the medium- and high-speed operating range, where model-based methods provide reliable and accurate performance. These approaches can be considered suitable for industrial applications. However, their limitations are most evident in the low-speed operating area, where the EMF is too small and conventional estimation techniques lose robustness and accuracy. Reliable position estimation at low and zero speeds therefore remains an open and important problem. In this region, saliency-based HF signal injection methods represent an effective solution because they use the magnetic anisotropy of the machine and enable position estimation. Square-wave injection allows higher injection frequencies but can introduce additional losses. The HF pulse-wave signal injection method combines structural simplicity with effective rotor position estimation. Therefore, future research will focus on its implementation using simulation and experimental validation. The objective is to assess its robustness, dynamic performance, and suitability for low- and zero-speed operation, and to compare pulse-wave signal injection with other HF injection methods.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under the project VEGA 1/0584/24 and by the Slovak Research and Development Agency under the contract No. APVV-23-0521.

References

[1] G. Wang, G. Zhang, and D. Xu, *Position Sensorless Control Techniques for Permanent Magnet Synchronous Machine Drives*, ser. Engineering. Springer Singapore, 2020, no. 1. [Online]. Available: https://doi.org/10.1007/978-981-15-0050-3

[2] V. Petro, "Bezsnímačové riadenie synchrónneho motora s permanentnými magnetmi pre nízke rýchlosti," Dissertation thesis, Technical University of Košice, Košice, Slovakia, 2023.

[3] G. Wang, M. Valla, and J. Solsona, "Position sensorless permanent magnet synchronous machine drives—a review," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5830–5842, 2020.

[4] Z. Q. Zhu and X. M. Wu, *Sensorless Control of Permanent Magnet Synchronous Machine Drives*, ser. Wiley-IEEE Press Series on Control Systems Theory and Applications. Wiley-IEEE Press, 2023, no. 1. [Online]. Available: https://doi.org/10.1002/9781394194384

[5] J.-M. Kim, S.-J. Kang, and S.-K. Sul, "Vector control of interior permanent magnet synchronous motor without a shaft sensor," in *Proceedings of APEC 97 - Applied Power Electronics Conference*, vol. 2, 1997, pp. 743–748 vol.2.

[6] J. Agrawal and S. Bodkhe, "Experimental study of low speed sensorless control of pmsm drive using high frequency signal injection," *Advances in Electrical and Electronic Engineering*, vol. 14, no. 1, pp. 29–39, 2016. [Online]. Available: https://advances.vsb.cz/index.php/AEEE/article/view/1564

[7] L. Jin, X. Wang, L. Lu, Y. Mao, and Z. Wang, "High-precision rotor position estimation via cascaded model-based and data-driven estimator: An integrated method," *IEEE Industrial Electronics Magazine*, vol. 19, no. 3, pp. 6–19, 2025.

[8] S. S and F. Muhammed, "Review of speed controlling techniques on dual pmsm drive systems," in *2022 International Conference on Futuristic Technologies in Control Systems Renewable Energy (ICFCR)*, 2022, pp. 1–6.

[9] J. Agrawal and S. Bodkhe, "Low speed sensorless control of pmsm drive using high frequency signal injection," in *2015 Annual IEEE India Conference (INDICON)*, 2015, pp. 1–6.

[10] G. Zhang, G. Wang, and D. Xu, "Saliency-based position sensorless control methods for pmsm drives - a review," *Chinese Journal of Electrical Engineering*, vol. 3, no. 2, pp. 14–23, 2017.

[11] S. Xu, D. Wang, B. Wang, X. Wang, W. Miao, and X. Wang, "Sideband electromagnetic vibrations in interior pmsm with high frequency pulsating injection sensorless control," *IEEE Transactions on Energy Conversion*, vol. 39, no. 3, pp. 1919–1929, 2024.

[12] L. Zhu, B. Li, G. Zhang, R. Xiang, H. Zhang, G. Wang, and D. Xu, "Pseudo-random frequency pulse voltage injection for sensorless ipmsm drives at low speeds," in *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, 2022, pp. 1–6.

[13] J. Cai, Y. Gu, A. David Cheok, and Y. Yan, "A survey of phase-locked loop technologies in sensorless position estimation of permanent magnet synchronous motor drives," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–16, 2024.

[14] P. Jansen and R. Lorenz, "Transducerless position and velocity estimation in induction and salient ac machines," *IEEE Transactions on Industry Applications*, vol. 31, no. 2, pp. 240–247, 1995.

[15] M. Corley and R. Lorenz, "Rotor position and velocity estimation for a salient-pole permanent magnet synchronous machine at standstill and high speeds," *IEEE Transactions on Industry Applications*, vol. 34, no. 4, pp. 784–789, 1998.

[16] J. M. Liu and Z. Q. Zhu, "Sensorless control strategy by square-waveform high-frequency pulsating signal injection into stationary reference frame," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 2, no. 2, pp. 171–180, 2014.

[17] R. Ni, D. Xu, F. Blaabjerg, K. Lu, G. Wang, and G. Zhang, "Square-wave voltage injection algorithm for pmsm position sensorless control with high robustness to voltage errors," *IEEE Transactions on Power Electronics*, vol. 32, no. 7, pp. 5425–5437, 2017.

[18] G. Xie, K. Lu, S. K. Dwivedi, J. R. Rosholm, and F. Blaabjerg, "Minimum-voltage vector injection method for sensorless control of pmsm for low-speed operations," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1785–1794, 2016.

Switching Field and Magnetoelastic Properties of Amorphous Glass-Coated Microwires

[1,2]*Samuel ONUFER (1st year)*
Supervisor: [2]Ladislav Galdun

[1]Dept. of Physics, FEI, Technical University of Košice, Slovak Republic

[2]RVmagnetics, a. s., Štúrova 27, 04001 Košice, Slovakia • EU

[1]samuel.onufer@tuke.sk, [2]galdun@rvmagnetics.com

***Abstract*—Amorphous glass-coated microwires represent a specific class of cylindrical soft magnetic materials whose magnetic behavior is strongly influenced by internal mechanical stresses generated during fabrication. In Fe-rich compositions with positive magnetostriction, these stresses lead to axial magnetic anisotropy and magnetic bistability characterized by abrupt magnetization reversal through a single large Barkhausen jump. The switching field is highly sensitive to mechanical stress, temperature, and excitation frequency, which enables its use in various sensing applications. This review summarizes the fabrication process, the magnetoelastic origin of anisotropy, the physical meaning of the switching field, its temperature dependence, and commonly used measurement techniques. Open research challenges related to stress modeling, thermal stability, and material optimization are also discussed.**

***Keywords*—amorphous magnetics microwires, magnetoelastic anisotropy, magnetic bistability, switching field, sensor.**

I. INTRODUCTION

Glass-coated magnetic microwires belong to a specific group of soft magnetic materials. They consist of very thin metallic wires covered by a thin glass layer (Fig. 1). Their magnetic properties are strongly influenced by internal mechanical stresses. These stresses arise mainly due to the difference in thermal expansion coefficients between the metallic core and the glass coating during the fabrication process [1], [2], [3], [4].

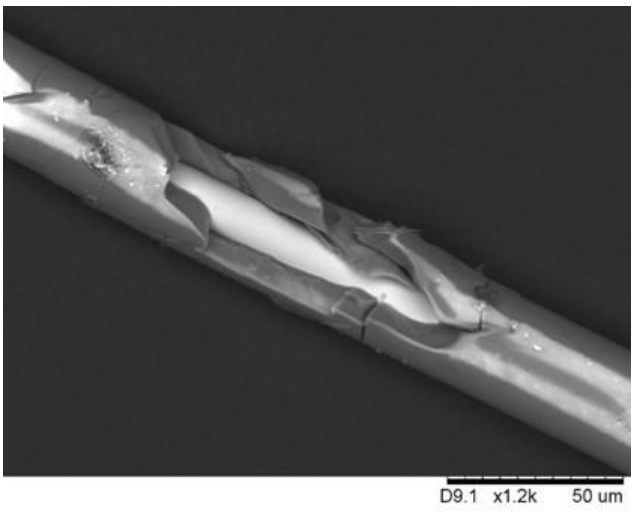


Fig. 1. SEM image of a glass-coated microwire showing the metallic core and the glass coating.

Unlike conventional crystalline magnetic materials, where magnetocrystalline anisotropy plays a dominant role, in amorphous microwires the key contribution comes from magnetoelastic coupling between magnetostriction and internal stress [5], [6]. This coupling determines the preferred direction of magnetization and, consequently, the overall magnetic behavior of the material.

In Fe-based microwires with positive magnetostriction, a specific domain structure is formed. The magnetization can exist only in two stable states. This phenomenon is referred to as magnetic bistability [6], [7]. Magnetization reversal occurs abruptly at a certain value of the external magnetic field, known as the switching field H_{sw}.

The switching field responds to various physical quantities, such as mechanical stress, temperature, position, and excitation frequency [8], [9], [10]. Due to this sensitivity, glass-coated microwires can be used as magnetic sensors in different applications [11].

II. CURRENT STATE OF RESEARCH

A. Fabrication and internal stress origin

Glass-coated microwires are produced using the Taylor–Ulitovsky process [1], [2], [3], [4]. In this technological method, a metallic pre-alloy is melted inside a glass capillary and then drawn together with the softened glass into a very thin filament (Fig. 2). During drawing, the metallic core is rapidly cooled, which prevents crystallization and results in an amorphous structure [3], [4].

The final structure consists of a metallic core covered by a thin glass layer. The process is very fast and allows the production of long segments of microwire with relatively small diameters. Rapid cooling is essential for the formation of the amorphous structure, which lacks long-range atomic order [5], [12].

During cooling, internal mechanical stresses are generated due to the difference in thermal expansion coefficients between the metallic core and the glass coating [3], [4]. The metal usually has a higher thermal expansion coefficient than glass. Therefore, during temperature decrease, the metal tends to contract more than the glass. The glass sheath restricts this contraction, which leads to the formation of residual stress in the metallic core.

As a result, axial tensile stress is mainly present in the core, while radial and circumferential stress components appear

near the surface. The stress distribution is not uniform across the cross-section of the wire. It is often described using a core–shell model, in which the inner part of the core and the region near the surface experience different stress conditions [3], [8].

These internal stresses strongly influence the magnetic behavior of the microwire. Through magnetoelastic coupling, mechanical stress is transformed into effective magnetic anisotropy, which determines the preferred direction of magnetization [5], [6]. For this reason, the fabrication process is directly linked to the resulting magnetic properties and cannot be considered separately.

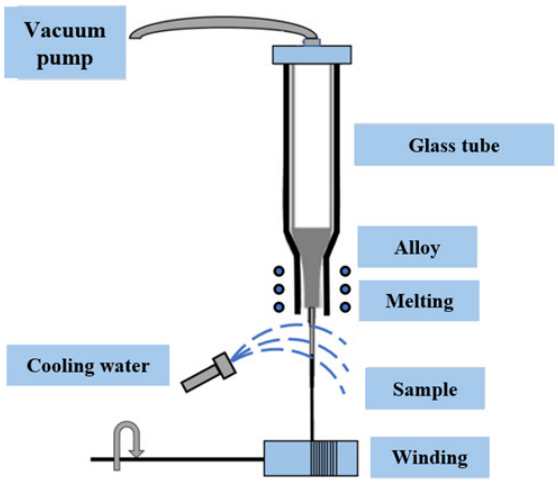


Fig. 2. Schematic illustration of the Taylor–Ulitovsky method for fabrication of glass-coated magnetic microwires [13].

In recent years, research has focused on modifying fabrication parameters and applying post-processing heat treatments to control the magnitude and distribution of internal stresses [14], [15], [16]. Even small changes in technological conditions can significantly affect the switching field and the stress sensitivity of the microwire.

B. Magnetoelastic anisotropy and bistability

In amorphous microwires, magnetocrystalline anisotropy is practically negligible, since the material does not possess a regular crystalline structure. Therefore, the dominant contribution to magnetic anisotropy is magnetoelastic anisotropy, which arises from the interaction between mechanical stress and a material with non-zero magnetostriction [5], [6], [17].

The magnetoelastic energy density can be expressed as:

$$E_{me} = -\frac{3}{2}\,\lambda_s\,\sigma\cos^2\theta,$$

where $\lambda_s$represents the saturation magnetostriction, $\sigma$is the effective mechanical stress, and $\theta$is the angle between the direction of magnetization and the direction of applied stress.

From this relation, it follows that magnetization tends to orient in a direction that minimizes magnetoelastic energy. In Fe-rich microwires with positive magnetostriction and axial tensile stress, it is energetically favorable for magnetization to align along the wire axis [6], [8].

As a result, a specific domain structure is formed. A large axial domain appears in the central region of the metallic core, while radial or circumferential magnetization components may occur near the surface (Fig. 3). Closure domains are formed at the wire ends in order to reduce magnetostatic energy associated with stray magnetic fields [6], [12].

This configuration leads to magnetic bistability. The magnetization in the axial core can exist only in two stable states – positive or negative along the wire axis. When the external magnetic field changes, magnetization does not rotate gradually but switches abruptly. This process occurs through domain wall propagation along the entire length of the microwire [7], [18].

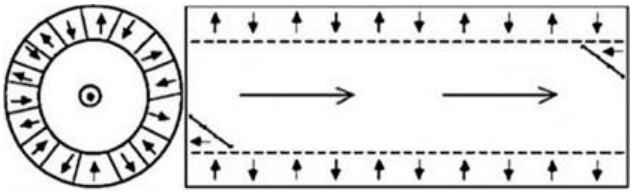


Fig. 3. Domain configuration and axial domain wall propagation in a bistable glass-coated microwire.

The switching field at which this transition occurs depends mainly on the magnitude of magnetoelastic anisotropy, that is, on magnetostriction and internal stress. In addition, local defects and structural inhomogeneities in the amorphous material can act as pinning sites for the domain wall and influence the switching behavior [7], [17], [19].

C. Switching Field and its Dependence

As mentioned above, magnetic bistability in Fe-rich microwires means that the magnetization of the core can exist only in two stable states. The transition between these states does not occur gradually but abruptly, at a certain value of the external magnetic field (Fig. 4). This value is referred to as the switching field H_{sw}.

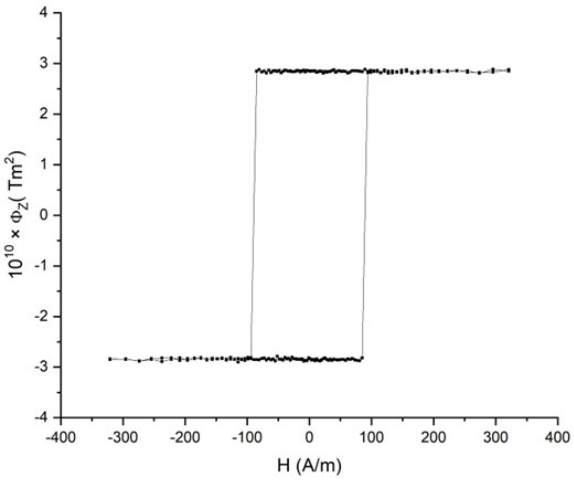


Fig. 4. Magnetic hysteresis loop of a bistable glass-coated microwire showing abrupt magnetization reversal at the switching field.

From a physical point of view, the switching field is related to the energy required to release the closure domain and to drive the domain wall along the microwire. Its magnitude is mainly influenced by magnetoelastic anisotropy, that is, by the combination of magnetostriction and internal stress.

In a simplified model, the switching field can be expressed as proportional to the ratio of magnetostriction and effective stress to the saturation magnetization:

$$H_{sw} \sim \frac{\lambda_s\,\sigma}{M_s}.$$

This relation shows that any change in internal or external stress leads to a change in the value of H_{sw}. Since magnetoelastic anisotropy depends on several physical quantities, such as mechanical stress, temperature, and excitation frequency, the switching field becomes a sensitive parameter that can be used to measure various physical quantities.

1) Dependence on Mechanical Stress

If additional axial tensile stress is applied to the microwire, the effective magnetoelastic anisotropy increases. In the case of positive magnetostriction, this further supports axial alignment of magnetization. As a consequence, the energy barrier that the domain wall must overcome increases, which leads to an increase in the switching field.

On the other hand, when compressive stress is applied, the effective anisotropy may decrease, resulting in a reduction of H_{sw}. Experimental results reported in several studies confirm an approximately linear dependence of the switching field on applied tensile stress within a certain range of loading [10], [20].

However, at higher stress levels, deviations from linear behavior may occur. This is mainly due to changes in domain structure or non-uniform stress distribution across the cross-section of the microwire.

2) Dependence on Temperature

The temperature dependence of the switching field can be significantly influenced by the presence of a nanocrystalline phase in the amorphous matrix. In FeMoBCu-based materials, suitable heat treatment leads to the formation of very fine crystalline grains dispersed in the amorphous phase [14], [15], [21].

If the grain size is smaller than the exchange length, magnetic moments in neighboring grains strongly interact and the effective anisotropy is averaged. The material then behaves magnetically soft, despite containing a crystalline phase [14], [15], [22].

However, at temperatures close to the Curie temperature of the amorphous matrix, exchange coupling between grains may weaken. Crystalline grains can then behave as individual magnetic regions and may act as obstacles to domain wall motion. In such cases, the switching field may increase with increasing temperature, as observed experimentally in some systems [9], [14], [22].

In contrast, in unannealed or predominantly amorphous materials, increasing temperature reduces effective anisotropy due to thermal fluctuations of magnetic moments. This leads to a decrease in the switching field [9].

From the point of view of sensor applications, it is therefore possible to tailor the temperature dependence of the switching field by selecting appropriate chemical composition and heat treatment conditions.

3) Dependence on Excitation Frequency

The switching field is not purely a static quantity. Its value also depends on the frequency of magnetic excitation. At low frequencies, relaxation processes have sufficient time to influence domain wall motion, which may lead to a different value of H_{sw}compared to higher frequencies [9].

At higher frequencies, the instantaneous magnetoelastic contribution dominates and the influence of relaxation is reduced. From a practical point of view, this means that excitation conditions must always be specified when measuring the switching field.

This frequency dependence also allows optimization of sensor performance. By selecting an appropriate excitation frequency, it is possible either to increase temperature sensitivity or to minimize temperature influence when measuring mechanical stress.

D. Experimental Methods for Microwire Characterization

When studying the properties of glass-coated microwires, it is important to select an appropriate measurement method. Due to the small dimensions of the metallic core and the specific domain structure, classical techniques used for bulk magnetic materials cannot always be applied directly.

1) Inductive Measurement of the Hysteresis Loop

The most commonly used method is inductive measurement. The microwire is placed inside an excitation coil that generates a time-varying magnetic field. The change in magnetization induces a voltage in a pickup coil, which can be recorded and analyzed [6], [8].

In bistable microwires, the hysteresis loop has an almost rectangular shape. Magnetization reversal does not occur gradually but appears as a single sharp voltage pulse corresponding to a Barkhausen jump [7], [18]. This pulse can be used to determine the switching field.

Since the metallic core has a very small volume, the induced voltage is relatively weak. Therefore, appropriate signal amplification and reduction of electromagnetic noise are necessary.

2) Switching Field Measurement Using the Time Method

In practice, the switching field is often not measured directly from the magnetic field amplitude but using a time-based method. The microwire is excited by a triangular magnetic field with a known slope. The moment at which magnetization switches is detected as a time marker.

Because the magnetic field changes linearly with time, the switching field can be calculated from the switching time. This method is particularly suitable for digital applications, since it allows simple signal processing using a microcontroller [10], [23].

It is important to ensure that the mechanical fixation of the microwire during measurement does not introduce additional stress that could influence the results.

III. OPEN PROBLEMS AND FUTURE DIRECTIONS

Although research on glass-coated microwires has been ongoing for several decades, several open questions still require further investigation.

As part of my doctoral research, I will specifically focus on optimizing the annealing process for Fe-Mo-B-Cu based microwires to achieve the highest possible sensitivity for sensor applications. I will investigate how different heat treatment conditions and the resulting structural relaxation influence the switching field. Furthermore, I will study the influence of the geometric parameters, specifically the diameter of the metallic core, the thickness of the glass

coating, and their ratio on the resulting sensor sensitivity. This research will also examine how these geometric factors affect the thermal range, identifying the maximum and minimum temperatures at which the microwire maintains its bistable behavior.

One of the main challenges is the accurate determination of internal stress distribution within the metallic core. Most existing models assume a simplified core–shell stress distribution. However, stress profile may be more complex and strongly dependent on specific fabrication conditions. More precise experimental techniques for stress characterization could significantly improve the understanding of magnetic behavior in microwires.

Another important issue is the long-term stability of the switching field. For practical sensor applications, it is essential that the material properties remain stable over time, for example under repeated mechanical loading or temperature cycling. Structural relaxation processes in amorphous materials may lead to gradual changes in magnetic properties, and this effect should be systematically studied.

The influence of decreasing core diameter is also an important topic. At very small dimensions, both the domain structure and the dynamics of domain wall motion may change. This can affect the sensitivity and stability of the switching field. Research in this area is particularly relevant for applications where sensor miniaturization is required.

Finally, standardization of measurement conditions remains a significant issue. The value of the switching field depends on excitation frequency, waveform shape, and mechanical fixation of the microwire. Without clearly defined experimental conditions, comparison of results obtained at different laboratories becomes difficult.

IV. Conclusion

In summary, glass-coated microwires represent a promising platform for stress and temperature sensing due to their magnetoelastic behavior and switching field sensitivity. Continued research focused on stress modeling, thermal stability and measurement standardization will further enhance their practical applicability.

Acknowledgment

This research was supported by VEGA Grant No. 1/0350/24 from the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic.

References

[1] G. F. Taylor, “A method of drawing metallic filaments,” *Physical Review*, vol. 23, pp. 655–660, 1924.

[2] A. Zhukov, J. González, J. M. Blanco, M. Vázquez, and V. Larin, “Microwires coated by glass: A new family of soft and hard magnetic materials,” *Journal of Materials Research*, vol. 15, no. 10, pp. 2107–2113, Oct. 2000. doi: 10.1557/jmr.2000.0303.

[3] V. S. Larin, A. V. Torcunov, A. Zhukov, J. González, M. Vázquez, and L. Panina, “Preparation and properties of glass-coated microwires,” *Journal of Magnetism and Magnetic Materials*, vol. 249, no. 1–2, pp. 39–45, Aug. 2002. doi: 10.1016/S0304-8853(02)00501-2.

[4] S. Baranov, V. Larin, and A. Torcunov, “Technology, preparation and properties of the cast glass-coated magnetic microwires,” *Crystals*, vol. 7, no. 6, p. 136, May 2017. doi: 10.3390/cryst7060136.

[5] J. Alam, C. Bran, H. Chiriac, N. Lupu, T. A. Óvári, L. V. Panina, V. Rodionova, R. Varga, M. Vazquez, and A. Zhukov, “Cylindrical micro and nanowires: Fabrication, properties and applications,” *Journal of Magnetism and Magnetic Materials*, vol. 513, p. 167074, 2020. doi: 10.1016/j.jmmm.2020.167074.

[6] P. Corte-León, A. Zhukov, M. Ipatov, J. M. Blanco, and V. Zhukova, “Magnetic microwires with unique combination of magnetic properties suitable for sensor applications,” *Sensors*, vol. 20, no. 24, p. 7203, Dec. 2020. doi: 10.3390/s20247203.

[7] R. Varga, “Switching field fluctuations in bistable microwires,” *Physica B: Condensed Matter*, vol. 343, no. 1–4, pp. 403–409, 2004. doi: 10.1016/j.physb.2003.08.077.

[8] A. Zhukov, M. Ipatov, P. Corte-León, L. Gonzalez-Legarreta, J. M. Blanco, and V. Zhukova, “Soft magnetic microwires for sensor applications,” *Journal of Magnetism and Magnetic Materials*, vol. 498, p. 166180, Mar. 2020. doi: 10.1016/j.jmmm.2019.166180.

[9] R. Sabol, R. Varga, J. Hudak, J. Blazek, D. Praslicka, P. Vojtanik, G. Badini, and M. Vazquez, “Temperature and frequency dependencies of the switching field in glass-coated FeNbSiB microwires,” *Journal of Applied Physics*, vol. 111, no. 5, 053919, 2012. doi: 10.1063/1.3691961.

[10] R. Sabol, M. Rovnak, P. Klein, M. Vazquez, and R. Varga, “Mechanical stress dependence of the switching field in amorphous microwires,” *IEEE Transactions on Magnetics*, vol. 51, no. 1, 2015. doi: 10.1109/TMAG.2014.2357580.

[11] V. Zhukova, J. M. Blanco, M. Ipatov, L. Gonzalez-Legarreta, S. Taskaev, and A. Zhukov, “Development of magnetic microwires for magnetic sensor applications,” *Sensors*, vol. 19, no. 21, p. 4767, Nov. 2019. doi: 10.3390/s19214767.

[12] H. Chiriac and T. A. Óvári, “Amorphous glass-covered magnetic wires: Preparation, properties and applications,” *Progress in Materials Science*, vol. 40, pp. 333–407, 1996.

[13] PARKHACHEV, V. N. Installation for production of glass insulated microwire directly from liquid metal. (US patent 3,256,584).

[14] J. Onufer, J. Ziman, P. Duranka, and M. Kladivová, “The influence of annealing on domain wall propagation in bistable amorphous microwire with unidirectional effect,” *Physica B: Condensed Matter*, vol. 540, pp. 58–64, Jul. 2018. doi: 10.1016/j.physb.2018.04.011.

[15] P. Corte-León, A. Zhukov, J. M. Blanco, M. Ipatov, and V. Zhukova, “Engineering of magnetic softness and domain wall dynamics of Fe-rich amorphous microwires by stress-induced magnetic anisotropy,” *Scientific Reports*, vol. 9, no. 1, 2019. doi: 10.1038/s41598-019-48755-4.

[16] V. Zhukova et al., “Optimization of magnetic properties of magnetic microwires by post-processing,” *Processes*, vol. 8, no. 8, 2020. doi: 10.3390/pr8081006.

[17] V. Zhukova et al., “The magnetostriction of amorphous magnetic microwires: The role of the local atomic environment and internal stresses relaxation,” *Magnetochemistry*, vol. 9, no. 10, 2023. doi: 10.3390/magnetochemistry9100222.

[18] R. Varga, K. L. Garcia, M. Vazquez, A. Zhukov, and P. Vojtanik, “Switching-field distribution in amorphous magnetic bistable microwires,” *Physical Review B*, vol. 70, 024402, 2004.

[19] K. L. Garcia, R. Varga, and M. Vazquez, “Influence of the magnetoelastic mechanism on the fluctuating switching field of Fe-based amorphous microwires,” *IEEE Transactions on Magnetics*, vol. 41, no. 10, pp. 3256–3258, Oct. 2005. doi: 10.1109/TMAG.2005.854669.

[20] M. A. Ali et al., “Application of bistable glass-coated microwire for monitoring and measuring the deformations of metal structural members,” *Measurement*, vol. 208, 112458, 2023. doi: 10.1016/j.measurement.2023.112458.

[21] V. Zhukova, J. M. Blanco, M. Ipatov, M. Churyukanova, S. Taskaev, and A. Zhukov, “Tailoring of magnetoimpedance effect and magnetic softness of Fe-rich glass-coated microwires by stress-annealing,” *Scientific Reports*, vol. 8, no. 1, 2018. doi: 10.1038/s41598-018-21356-3.

[22] V. Zhukova, M. Churyukanova, S. Kaloshkin, P. Corte-León, M. Ipatov, and A. Zhukov, “Magnetostriction of amorphous Co-based and Fe-Ni-based magnetic microwires: Effect of stresses and annealing,” *Journal of Alloys and Compounds*, 2023. doi: 10.1016/j.jallcom.2023.170122.

[23] P. Jacko, P. Duranka, and R. Varga, “Advantages of bistable microwires in digital signal processing,” *Sensors*, vol. 24, no. 8, 2024. doi: 10.3390/s24082423.

Analysis and Modeling of Electricity Market Dynamics under Increasing Renewable Integration

[1]*František KURIMSKÝ (2nd year)*
Supervisor: [2]Marek PAVLÍK

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]frantisek.kurimsky@tuke.sk, [2]marek.pavlik@tuke.sk

***Abstract*—European electricity markets are experiencing significant structural transformation due to the rapid integration of renewable energy sources. This research analyzes the evolving dynamics of electricity prices under increasing renewable penetration, focusing on price volatility, renewable market value, and structural price formation patterns. The study examines long-term trends across selected European markets and evaluates how renewable production levels influence price behavior under different market regimes, including periods of extreme volatility. The findings provide deeper insight into renewable-driven price dynamics and contribute to improved understanding of market stability, investment signals, and future forecasting challenges.**

***Keywords*—Electricity market, Electricity price, Photovoltaic and Wind generation, Predictions, Renewable integration**

I. Introduction

Over the last years, European electricity market has undergone transformation driven primarily by the accelerated integration of renewable energy sources. This development is strongly linked to climate objectives of European Union, particularly the 2050 climate neutrality target, the European Green Deal [1], and legislative frameworks such as Renewable Energy Directive (RED III), that raises the binding targets for renewable energy and updates rules to accelerate the deployment of renewables across sectors to help meet climate and energy goals by 2030 [2].

At the same time, geopolitical tensions such as war on Ukraine and the Covid-19 pandemic altered market behavior. These external shocks exposed structural vulnerabilities of the European power system and significantly increased price volatility. Day-ahead market prices became more unstable, with frequent price spikes and an unprecedented growth in negative price occurrences.

The increasing share of variable renewable generation has introduced new dynamics into electricity price formation. While renewables reduce average marginal costs through the merit-order effect, they also increase short-term variability due to their weather-dependent nature.

This research analyzes electricity market dynamics under increasing renewable penetration and develops data-driven analytics that supports deeper understanding of price behavior, volatility, renewable market value, and future forecasting applications.

II. Background

The European electricity market operates through interconnected energy exchanges. Electricity trading can be divided based on time horizons into:

- Long-term (futures contracts),
- Day-Ahead Market (DAM),
- Intraday Market (IDM),
- Balancing Market.

In the DAM, participants submit bids representing specific market orders, representing volume and price of traded electricity [3]. IDM trading allows adjustment closer to real time [4], [5]. Balancing markets ensure system stability, when deviations between planned and actual production and demand occur. Market participants who fail to meet their committed schedules face imbalance settlement costs.

A. Price Formation Mechanism

Electricity price formation for DAM market and auctions on IDM is based on EUPHEMIA algorithm [6]. This algorithm determines the market clearing price (MCP) for each bidding zone while maximizing economical effectiveness of whole system and respecting network contains. Although this algorithm ensures economic efficiency at a theoretical level, increasing renewable penetration introduces challenges, for example, high RES output may cause local congestion [7]. Generation from renewable sources across neighboring regions also limit export possibilities, due to corelated weather condition, which causes price depression in interconnected markets [8]. In addition, when renewable production exceeds demand and system flexibility is insufficient, negative prices may occur more frequently. These situations indicate that increasing renewable penetration fundamentally reshapes electricity market dynamics and price behavior.

III. Recent Achievements

During the last academic year, we developed and published several analytical and modeling contributions focused on European electricity markets.

In [9] we analyzed evolution of electricity price volatility in six European countries over the years 2015-2025, focusing

on the impact of increasing renewable energy penetration. The results show a significant increase in price volatility after 2021, particularly during the 2022 energy crisis, with the strongest effects observed in countries with a high share of wind and solar generation and limited system flexibility. The study identifies non-linear threshold effects, where the probability of negative prices increases sharply once renewable generation exceeds certain levels. The article concludes that while renewable energy reduces average electricity prices in the long term due to the merit-order effect, it increases short-term price volatility and the frequency of extreme price events. Therefore, further development of system flexibility, energy storage, demand response, and cross-border integration is essential to maintain market stability under high renewable penetration.

In [10] we applied Benford's law to the absolute hourly differences of DAM electricity prices in six European countries over the period 2015-2024. We examined whether electricity price variations follow the expected digit distribution define by the Benford's law and whether events such as Covid-19 pandemic and the war in Ukraine influenced this distribution. We performed analysis for the first digit distribution (BL1), second digit distribution (BL2) and the first 2 digits distribution (BL12). The results show very strong conformity with law across all countries, with different energetical mixes, confirmed by chi-square tests with p-values close to 1. Slightly larger deviations from law were observed for the BL2 in some years, but no systematic irregularities were detected. The external shocks mentioned above did not cause deviations from the expected digit distribution. The study concludes that Benford's Law can serve as a useful statistical tool for anomaly detection, market monitoring, and identifying potential irregularities or manipulations in electricity markets.

In our previous research [11] we analyzed the capture price (CP) and capture rate (CR) for photovoltaic power plants in selected Central European countries, focusing on how volatile electricity prices influence the economic performance of solar energy. In the present study "submitted for publication" [12], we continue this research framework and extend the CP and CR methodology to wind power generation over the period 2015-2025, allowing for a long-term assessment of wind's market value under different price regimes, including the 2021-2022 energy crisis. By comparing Germany and Poland, we evaluate how renewable penetration levels influence wind capture performance and identify structural differences in market realization. In Germany, capture prices are consistently below average day-ahead prices in most months, particularly during winter and crisis years, leading to capture rates frequently below 90% and reaching a minimum of 61% in December 2022. In contrast, Poland exhibits significantly more stable capture dynamics. Capture prices closely follow average market prices, and capture rates generally fluctuate between 90% and 100%, with several months exceeding 100%. Even during the extreme volatility of 2022, Poland's capture rate remained relatively resilient compared to Germany. In 2025 the situation shows partial improvement, with capture rates recovering in several months. These findings suggest that wind market value decline is strongly dependent on penetration level and system flexibility, rather than being a universal phenomenon across European markets.

In the present study "to be published" [13], we extend the analysis of renewable market dynamics by applying The General Unary Hypotheses Automaton (GUHA) association rule mining framework to identify interpretable relationships between solar and wind generation levels and day-ahead electricity price categories in Germany (2015-2024). Unlike regression-based or forecasting approaches, GUHA enables hypothesis-free discovery of statistically significant rules, allowing us to detect conditional and time-specific patterns in complex market data. The analysis was implemented using the CleverMiner 4ft-Miner tool, applied only to hours with solar generation of at least 50 MWh, with a minimum confidence threshold of 0.6 and minimum support of 30, and antecedent length constrained to 3-5 attributes from solar, wind onshore, wind offshore, year, and hour variables. Applying the GUHA method to hourly German electricity market data (2015–2024) allowed us to extract 248 statistically significant association rules linking solar and wind generation levels, temporal conditions (hour, year), and electricity price categories. Each extracted finding is expressed as an association rule of the form Antecedent (A) ⇒ Consequent (C), where the antecedent represents specific combinations of renewable generation categories, temporal variables, and year, and the consequent corresponds to a defined electricity price category.

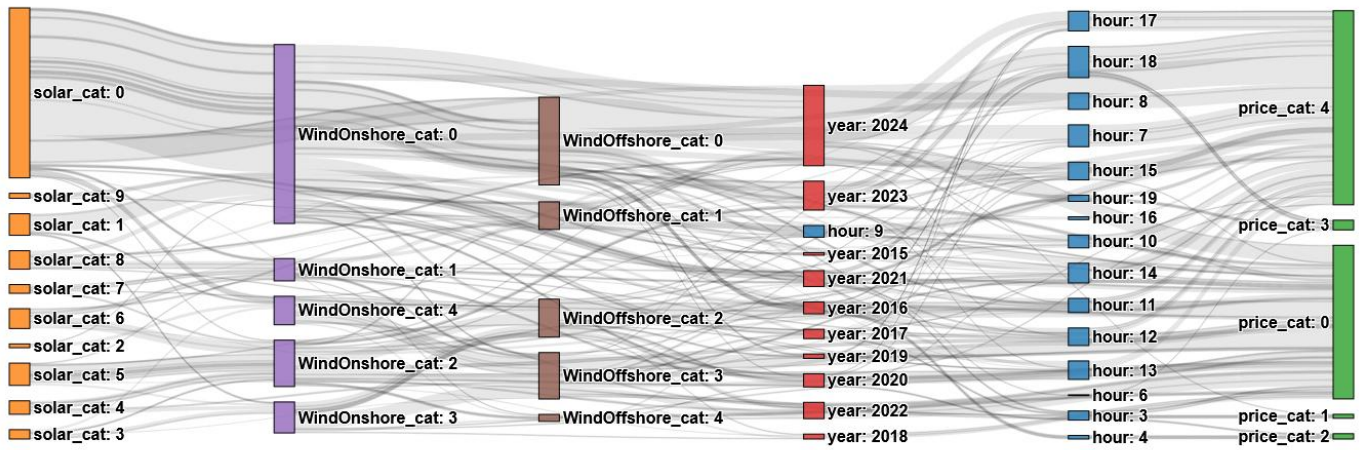


Fig. 1 Sankey diagram visualization of GUHA association rules showing the flow from solar and wind generation categories through temporal variables (year and hour) to electricity price categories in the German day-ahead market (2015–2024)

Combinations of high solar and moderate-to-high wind generation are systematically associated with the lowest price category, especially during midday hours, while low renewable generation, particularly in evening peak hours and in recent years such as 2024, are strongly linked to high price outcomes. Using a Python-based Sankey diagram library, we recently developed a multi-layer flow representation, presented in Fig. 1, where each layer represents one step in the association rule (solar, wind onshore, wind offshore, year, hour, price category), with each variable discretized into quantile-based intervals. Each pathway represents a statistically significant association rule, and the number of connections between nodes reflects the number of co-occurrences across extracted rules. The visualization clearly highlights dominant pathways, such as the strong flow from low solar and low wind production in 2024 evening hours toward the highest price category. This graphical representation enhances interpretability of the GUHA rules and provides an intuitive overview of structural market behavior beyond tabular rule outputs.

In conference paper [14] we developed deep learning models to estimate hourly onshore and offshore wind power generation in Germany. Wind speed components (u_{100} and v_{100}) from the Copernicus ERA5 dataset were combined with real wind generation and installed capacity data from Energy-Charts for the period 2019–2024, with testing performed on data from January to May 2025. Two architectures were compared, a fully connected regression model and a Convolutional Neural Network (CNN). Models used spatial wind field data and installed capacity as inputs and produced outputs for onshore and offshore generation. The CNN significantly outperformed the regression approach, achieving lower validation and test errors (MAE and MSE), demonstrating better ability to learn spatial wind patterns from ERA5 data. Onshore wind generation was predicted more accurately than offshore generation in both models, likely due to the limited spatial grid coverage for offshore wind areas. The study concludes that convolutional architectures are more suitable for wind power estimation with spatial meteorological data. Future research should include additional meteorological variables and explore hybrid architectures combining CNNs with temporal models (e.g., LSTM) to further improve forecasting accuracy.

Overall, the recent research activities demonstrate a comprehensive analytical framework combining statistical analysis, association rule mining, and deep learning methods - first to understand electricity market dynamics under increasing renewable penetration, and subsequently to develop generation and price forecasting models as core objectives of the dissertation.

IV. Future Plans

In the forthcoming research period, we plan to further improve renewable generation and electricity price forecasting by implementing systematic hyperparameter optimization for neural network architectures applied to solar and wind prediction. In parallel, the GUHA association rule mining framework will be extended to additional European countries to identify cross-country structural differences in renewable-price interactions. The GUHA analysis will also incorporate additional system variables, particularly load, residual load, and cross-border electricity flows, to capture more complex relationships influencing electricity price formation.

Furthermore, we intend to investigate the applicability of Temporal Fusion Transformer (TFT) architectures for electricity price prediction, evaluating whether their multi-horizon and temporal attention mechanisms can enhance modeling accuracy under volatile market conditions. These steps aim to strengthen both the interpretability and predictive capabilities of our research and will directly contribute to the forecasting part of the dissertation, completing the path from understanding electricity market dynamics toward accurate price and generation prediction under increasing renewable penetration.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00015.

References

[1] "The European Green Deal - European Commission." Accessed: Dec. 01, 2025. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

[2] *DIRECTIVE (EU) 2023/2413 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. 2023. Accessed: Feb. 16, 2026. [Online]. Available: http://data.europa.eu/eli/dir/2023/2413/oj

[3] "Single Day-Ahead Coupling (SDAC)." Accessed: Dec. 06, 2025. [Online]. Available: https://www.nemo-committee.eu/sdac

[4] "Intraday Auctions on SIDC." Accessed: Dec. 08, 2025. [Online]. Available: https://www.nemo-committee.eu/ida

[5] "Continuous trading." Accessed: Dec. 08, 2025. [Online]. Available: https://www.nemo-committee.eu/continuous

[6] "EUPHEMIA Public Description." Accessed: Feb. 24, 2025. [Online]. Available: https://nemo-committee.eu/assets/files/euphemia-public-description.pdf

[7] "More coordination needed in renewable deployment to prevent grid congestion - Joint Research Centre." Accessed: Feb. 16, 2026. [Online]. Available: https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/more-coordination-needed-renewable-deployment-prevent-grid-congestion-2024-05-27_en

[8] C. Stiewe, A. L. Xu, A. Eicke, and L. Hirth, "Cross-border cannibalization: Spillover effects of wind and solar energy on interconnected European electricity markets," May 27, 2024, *arXiv*: arXiv:2405.17166. doi: 10.48550/arXiv.2405.17166.

[9] M. Pavlík, F. Kurimský, and K. Ševc, "Renewable Energy and Price Stability: An Analysis of Volatility and Market Shifts in the European Electricity Sector (2015–2025)," *Applied Sciences*, vol. 15, no. 12, p. 6397, Jan. 2025, doi: 10.3390/app15126397.

[10] F. Kurimský, M. Pavlík, J. Petráš, and A. Hyseni, "The Application of Benford's Law on the Hourly Differences of Electricity Prices," *Acta Polytechnica Hungarica*, vol. 22, no. 11, 2025, doi: 10.12700/APH.22.11.2025.11.7.

[11] M. Pavlík, M. Bereš, and F. Kurimský, "Analyzing the Impact of Volatile Electricity Prices on Solar Energy Capture Rates in Central Europe: A Comparative Study," *Applied Sciences*, vol. 14, no. 15, Art. no. 15, Jan. 2024, doi: 10.3390/app14156396.

[12] M. Pavlík, F. Kurimský, and K. Ševc, "Long-Term Capture Rate Analysis of Wind Energy in Europe: Market Dynamics and Profitability," presented at the Advanced Power Systems 2026, Cluj-Napoca, Romania, Mar. 2026, **To be published**.

[13] F. Kurimský, M. Pavlík, and Ľ. Antoni, "Analyzing the Association between Solar and Wind Generation and Electricity Prices in Germany using GUHA Association Rules," presented at the IEEE 13th International Joint Conference on Cybernetics and Computational Cybernetics, Cyber-Medical Systems, Brisbane, Australia: IEEE Hungary Section, Apr. 2026, **To be published**.

[14] F. Kurimský and M. Pavlík, "Neural Network-Based Estimation of Onshore and Offshore Wind Power Generation in Germany," in *Kandó Konferencia 2025*, Budapest: Kandó Kálmán Villamosmérnöki Kar, Nov. 2025, pp. 374–384. [Online]. Available: https://konf.kvk.uni-obuda.hu/en/publications/

Large Language Models in Energy Commodity Trading

[1]*Alex KREMPASKÝ (1st year),*
Supervisor: [2]*Peter Papcun, Consultant:* [3]*Erik KAJÁTI*

[1,2,3]Institute of Artificial Intelligence, FEI, Technical University of Košice, Slovak Republic
[1]alex.krempasky@tuke.sk, [2]peter.papcun@tuke.sk, [3]erik.kajati@tuke.sk

***Abstract*—Global energy commodity markets span electricity, natural gas, coal, oil, and emissions permits, generating enormous annual turnover. Price volatility in these markets remains severe, driven by renewable energy variability, supply chain disruptions, and rapid shifts in market sentiment. Large Language Models (LLMs) augmented with advanced feature engineering, retrieval-augmented generation (RAG), and hybrid machine learning architectures demonstrate measurable improvements in price prediction accuracy compared to traditional statistical methods, and have already been used for multi-asset portfolio construction and sustainable energy decision support. However, deployment of LLM-based systems in real-world trading remains largely unexplored, constrained by regulatory compliance requirements, explainability mandates, and operational latency constraints that academic research typically overlooks. This paper analyzes peer-reviewed sources spanning energy markets, trading operations, carbon and electricity markets, and machine learning to identify the current state of LLM applications in commodity trading and to document critical barriers to production deployment.**

***Keywords*—energy trading, large language models, multi-commodity coupling, explainable AI, regulatory compliance, electricity price forecasting**

I. Introduction

Global energy markets generate trillions in annual turnover and remain critical to economic stability and the energy transition. Electricity, natural gas, and oil prices react rapidly to weather events, supply disruptions, policy announcements, and market sentiment—often captured first in news and regulatory filings rather than structured market data. Traditional forecasting approaches such as ARIMA and GARCH models capture autoregressive patterns and volatility clustering but struggle to integrate unstructured textual information containing crucial forward-looking signals [1], [3], [4], [24].

Large language models (GPT-4/5, Claude 4.5, Mistral, Llama) offer novel capabilities: they extract structured information from text, score sentiment, identify causal narratives, and generate feature representations for downstream predictors. Recent work shows LLMs as feature extractors, sentiment engines, or RAG-based modules improve electricity price predictions, portfolio diversification, and sustainable energy decision support [2], [9], [19], [20].

Energy trading operates under strict regulatory oversight (MiFID II, MAR, CFTC, GDPR) requiring audit trails, manipulation detection, position limits, and explainable rationales. Equally critical, markets are multi-commodity: traders hedge electricity with gas/oil futures and manage CO2 exposure, so trading decisions must account for cross-commodity risk rather than treating electricity in isolation. [6], [8], [7], [23].

This article analyzes LLM applications (FAEP, TimeXer+LLM, trading frameworks, carbon agents), identifies gaps in multi-commodity modeling and regulatory compliance, explains their limitations for regulated trading, and proposes research directions for explainable, regulation-ready energy trading systems [1], [3], [21], [23].

II. Current Approaches and State-of-the-Art

A. LLM-Enhanced Feature Engineering for Volatility Prediction: FAEP

The FAEP framework addresses electricity price volatility forecasting on the Australian New South Wales market by using an LLM with retrieval-augmented generation to automate feature engineering rather than to forecast prices directly. The system scans weather reports and incident documents to identify task-relevant factors (temperature, wind, demand), which are then compressed via Kernel PCA and fed into a hybrid XGBoost–LSTM model [1].

FAEP achieves a 32.18% reduction in MAE (from 0.87 to 0.59) relative to the HARQ-L-CJ benchmark, with RAG over historical reports, volatility decomposition, and price-jump features driving most of the gain. XGBoost captures nonlinear interactions and LSTM models temporal dependencies, illustrating that LLMs can significantly simplify and improve feature engineering for electricity price volatility [1].

For practical trading, however, FAEP has two major limitations. It ignores gas, oil, and CO_2 prices, so it cannot capture electricity–gas or electricity–CO_2 relationships that are central to hedging and spread trading, and it relies on Kernel PCA, which destroys interpretability. Because the paper does not describe any audit trail or explanation mechanism, deploying FAEP in a regulated environment would require substantial additional compliance and governance layers [1], [6], [8].

B. Sentiment-Augmented Price Forecasting: TimeXer+LLM

TimeXer+LLM combines LLM-derived news sentiment with a Transformer-based time-series model for CAISO day-ahead electricity prices. The framework processes 2,351 news articles, from which an LLM generates 48-dimensional impact vectors representing expected hourly price effects; these are aggregated into sentiment time series and used as exogenous inputs to the TimeXer model alongside prices, load, and renewable generation [3].

The model delivers a 2.6% MAE improvement at a 96-hour horizon, smaller gains at 144 hours, and degradation at

192 hours, indicating that sentiment is mainly a short-term signal. This shows that LLM-derived sentiment can be effectively integrated into transformer architectures without major redesign, but the approach still ignores gas and CO_2 coupling and uses an opaque 48-dimensional sentiment vector with no clear mapping to specific news themes, which complicates auditability and regulatory defense of trading decisions [3], [6].

C. LLM-Based Trading and Market Analytics in Electricity and Carbon Markets

Beyond FAEP and TimeXer+LLM, several frameworks bring LLMs closer to trading decisions and market supervision. An LLM-enhanced trading decision framework with multi-scale memory for the Australian spot market combines electricity prices with news text stored in short-, medium-, and long-term memory, achieving higher cumulative returns than deep reinforcement learning baselines in battery storage arbitrage and providing textual rationales for trades [21]. It improves adaptivity and interpretability but still focuses on a single commodity and does not address cross-commodity risk or explicit regulatory constraints.

In market surveillance, an AI-based market power calculation system for the Chinese electricity market uses a DeepSeek LLM as an interaction, reasoning, and RAG layer on top of Java-based numerical engines, computing HHI, top-m shares, and behavioral indicators via natural-language queries [22]. Similarly, generative-agent frameworks for carbon markets use LLM agents with RAG to integrate policy documents and market data, simulate stakeholder behavior in ETS systems, and improve carbon price forecasts [23]. These works illustrate the value of LLMs for decision support and supervision in electricity and CO_2 markets, but none provides an integrated, multi-commodity trading agent with full auditability [22], [21], [23].

D. Broader LLM Applications

Outside energy trading, LLM+RAG architectures have been successfully applied to multi-asset portfolio construction and sustainable energy decision support. In portfolio management, a RAG–Llama 3.1 system constructs diversified portfolios across equities, ETFs, commodities, cryptocurrencies, and bonds and outperforms vanilla LLM approaches in terms of return and risk [19].

In sustainable energy, an RAG–LLM QA assistant integrates regulatory and environmental data to help SMEs navigate subsidies, compliance, and net-zero initiatives, with large gains in precision and recall once retrieval is enabled [20].These examples suggest that the core LLM+RAG technology is mature; the remaining open questions in energy trading concern domain-specific multi-commodity coupling, regulatory constraints, and integration with physical and market limitations [11].

All three LLM-based frameworks in Table I improve forecasting or trading decisions for single-commodity electricity tasks, but none models multi-commodity coupling or provides full regulatory auditability. This reinforces the conclusion that current LLM applications in energy are valuable but incomplete from the perspective of regulated multi-commodity trading [1], [3], [21].

III. Multi-Commodity Coupling

The surveyed LLM-based frameworks for electricity forecasting and carbon-market analytics consistently operate on single-commodity setups and do not model electricity–gas–CO2 interactions explicitly.

A. Single-Commodity Models

Both FAEP and TimeXer+LLM treat electricity as an isolated forecasting target, and even the LLM-based multi-scale memory trading framework focuses on a single electricity market. Dataset availability favors such models, because public electricity price time series from CAISO, AEMO, or European TSOs are relatively easy to access, while building aligned multi-commodity datasets requires integrating proprietary natural gas, oil, and CO_2 series with consistent timestamps and units. Model complexity and publication incentives also favor single-commodity designs [1], [3], [21].

From a trading standpoint, this simplification is problematic. Electricity-only models cannot see gas supply shocks, oil-indexed LNG price moves, or CO_2 policy changes that propagate into power prices. They also ignore structural relationships between electricity and carbon markets that are increasingly important under net-zero policies and ETS expansions [7], [23].

B. Consequences for Trading

Ignoring multi-commodity coupling has several concrete consequences. Traders using single-commodity LLMs systematically misestimate portfolio risk: a model that forecasts flat electricity prices while gas prices spike due to a geopolitical shock will understate risk until electricity prices adjust, leading to unexpected losses. During recent European gas crises, electricity prices reacted strongly to gas supply constraints—behavior that electricity-only models could capture only ex post [7].

Single-commodity models also forgo spread-based arbitrage opportunities. If an LLM sees rising electricity prices but does not process gas data, it cannot distinguish between a gas-driven move, where electricity–gas spreads are likely to compress, and a generator outage, where spreads may remain wide. Profitable cross-commodity strategies rely precisely on these distinctions [1], [3].

Finally, portfolio completeness is an issue. Professional energy traders manage portfolios that include electricity, gas, oil, and CO_2 exposures, often hedged across regions and maturities. A forecasting system that covers only one leg of this structure is insufficient input for risk management and capital allocation decisions [18], [23].

C. Towards Multi-Commodity LLM Architectures

Addressing this gap requires architectures that ingest synchronized multi-commodity inputs, share textual understanding across commodities, and explicitly model cross-commodity attention. A natural design is to process electricity, natural gas, crude oil, and CO_2 allowance price histories in parallel; use a shared LLM text encoder for unified news and policy corpora; and add commodity-specific decoders with cross-commodity attention layers trained under a multi-task loss that reflects portfolio priorities [1], [3].

TABLE I
COMPARISON OF LLM-BASED APPROACHES IN ENERGY COMMODITY FORECASTING AND TRADING

Aspect	**FAEP[1]**	**TimeXer+LLM[3]**	**LLM Trading Framework[21]**
Primary Task	Electricity price volatility forecasting	Day-ahead electricity price forecasting	Trading decisions and battery arbitrage
LLM Role	RAG-based automated feature engineering from weather and incident reports	Sentiment extraction from news articles	Multi-scale memory integration and decision rationale generation
Architecture	LLM + RAG + Kernel PCA + XGBoost + LSTM	LLM sentiment encoder + TimeXer (Transformer)	LLM with short-, medium-, and long-term memory layers
Market/Dataset	Australian NSW electricity market	CAISO day-ahead market, 2,351 news articles	Australian electricity spot market
Key Results	32.18% MAE reduction (0.87 $\rightarrow$ 0.59) vs HARQ-L-CJ baseline	2.6% MAE improvement at 96h horizon; smaller gains at 144h	Higher cumulative returns vs deep RL baselines
Multi-Commodity	No (electricity only)	No (electricity only)	No (electricity only)
Explainability	Limited (Kernel PCA destroys interpretability)	Limited (opaque 48-dim sentiment vectors)	Partial (provides textual rationales but no quantitative attribution)
Regulatory Audit Trail	Not addressed	Not addressed	Not addressed
Cross-Commodity Coupling	Ignores gas, oil, CO_2	Ignores gas, oil, CO_2	Ignores gas, oil, CO_2

Evaluation should go beyond per-commodity MAE to include whether model-predicted correlations match empirical ones (e.g., winter electricity–gas correlation in Europe, oil–gas linkages in LNG-indexed contracts, electricity–CO_2 coupling in ETS regions), and whether simulated portfolios constructed from model outputs meet risk and drawdown constraints under stress scenarios [7], [23], [19].

IV. EXPLAINABILITY AND REGULATORY AUDITABILITY

Neither FAEP, TimeXer+LLM, nor current LLM-based electricity trading frameworks generate fully structured, regulator-ready explanations and audit trails. All rely on neural network components and opaque embeddings that are difficult to interpret, and logging strategies are often absent or ad hoc. For academic research, this is tolerable; for regulated trading, it is not [1], [3], [21].

A. Explainability Requirements

MiFID II, MAR, and GDPR impose explicit obligations on companies using algorithmic trading systems. They require detecting situations in which algorithms may enable market abuse, restricting automated decisions that produce legal or similarly significant effects, and providing meaningful information about the logic involved in automated decisions. Market surveillance regimes also demand reconstructing past trades with their underlying rationale [6], [8], [22].

Operationally, risk managers must understand why a model is long or short a given commodity, how it reacts to particular news themes, and under what conditions its recommendations should be overridden. Engineers need to diagnose failures, and compliance staff must be able to explain the system to regulators in non-technical language [22], [9].

B. Barriers in Current LLM Energy Forecasting

Three barriers recur across current work. Kernel PCA and similar nonlinear compression techniques destroy interpretability by mixing all inputs into opaque latent dimensions. High-dimensional sentiment or impact vectors produced by LLMs have no direct human-readable semantics, making it impossible to say which categories of news drove a particular score. Finally, logging is typically limited to predictions and realized errors, not to full input–feature–explanation tuples [1], [3].

Electricity trading frameworks with LLM-based reasoning do provide textual rationales in some cases, but these are not tied to quantitative explanation methods such as SHAP or attention visualization, and they are not systematically stored for audit. Market power supervision systems demonstrate how LLMs can be combined with RAG, tool calling, and structured prompts to explain indicator calculations (e.g., HHI, RSI), suggesting patterns that could be adapted to trading systems [21], [22].

C. Explainable and Auditable LLM Trading Systems

Explainable trading systems should avoid unnecessary opaque compression, maintain named features for key signals (e.g., gas-hedge indicator, CO_2-sensitivity feature, sentiment category scores), and exploit transformer attention maps and SHAP-style attributions to provide concise explanations for each decision. For example, a system may log that a buy recommendation in electricity was driven mainly by rising gas prices, bullish policy news, and low CO_2 prices [16], [9].

Structured audit logging must record timestamps, raw inputs (prices, order books if relevant, text snippets or sentiment scores), derived features, model versions and hyperparameters, decisions and position changes, and associated explanations. For high-risk trades, human-in-the-loop review with explanation-based gating can ensure that automated recommendations align with risk appetite and regulatory expectations [6], [8].

V. ENERGY COMMODITY TRADING REMAINS UNDEREXPLORED

While equity and crypto markets already have LLM trading benchmarks and agent arenas, energy commodity trading with LLMs remains substantially underexplored. Data access barriers, regulatory fragmentation, smaller commercial incentives, and the need to model physical constraints all contribute to this gap [11], [12], [13].

At the same time, related domains are progressing quickly. Agent-based RAG frameworks for carbon markets show how

LLM agents can combine policy retrieval, simulation, and causal reasoning to analyze ETS design and trading decisions [21], [23], [20].

What is missing is an integrated, multi-commodity, explainable, regulation-aware framework that extends these ideas to real-world portfolios of electricity, gas, oil, and CO_2, and that is evaluated under realistic trading constraints including settlement intervals, liquidity, and risk limits [1], [3], [7].

VI. Implications and Recommendations

For practitioners, single-commodity LLM forecasts should not be deployed as primary decision engines in multi-commodity portfolios; instead, they can serve as complementary signals within a broader risk framework that explicitly accounts for gas, oil, and CO_2 exposures. Companies should design explainability and auditability into their architectures from the start rather than attempting to retrofit them after optimizing pure accuracy metrics [1], [3], [6].

For researchers, promising directions include developing synchronized multi-commodity datasets with aligned textual corpora; designing LLM-based architectures that share textual understanding across commodities while allowing commodity-specific decoders; and integrating explanation mechanisms and logging schemas that are explicitly mapped to MiFID II, MAR, and GDPR requirements. Benchmarks for energy trading should evaluate both financial performance and compliance readiness [19], [23], [22].

VII. Conclusion

Large language models offer genuine promise for energy commodity trading by extracting meaning from unstructured text and enriching forecasting pipelines with signals that traditional time-series methods cannot access. Existing work on FAEP, TimeXer+LLM, electricity trading with multi-scale memory, carbon-market generative agents, and RAG-LLM decision-support systems shows that LLMs can already improve volatility prediction, sentiment integration, and contextual understanding in both energy and finance [1], [3], [4], [19], [20], [23].

However, current architectures remain far from deployable multi-commodity trading systems. They largely ignore the structural coupling between electricity, gas, oil, and CO_2 prices, and they do not provide the level of explainability, logging, and governance required by MiFID II, MAR, GDPR, and analogous regimes [6], [8], [7].

In future work, we plan to move from this qualitative gap analysis to a concrete, regulation-aware LLM+RAG prototype. The first step is to build a synchronized multi-commodity dataset for European electricity, gas, oil, and CO2, aligned with timestamped news and regulatory texts. On top of this, we will design a multi-commodity LLM+RAG architecture whose retrieval is conditioned on portfolio exposures and whose outputs include structured, regulator-ready explanations rather than opaque scores. The prototype will be backtested against existing single-commodity LLM frameworks, with evaluation criteria that combine financial performance, risk metrics, and the completeness of the audit trail.

Acknowledgment

This publication was supported by the VEGA grant GreenEDGE: Intelligent systems for sustainable symbiosis (1/0556/26).

References

[1] H. Xue *et al.*, "LLM-enhanced feature engineering for multi-factor electricity price predictions," arXiv preprint arXiv:2505.11890, May 2025. [Online]. Available: https://arxiv.org/abs/2505.11890

[2] S. Madani, A. Tavasoli, Z. Khoshtarash Astaneh, and P.-O. Pineau, "Large language models integration in smart grids," arXiv preprint arXiv:2504.09059, Apr. 2025. [Online]. Available: https://arxiv.org/abs/2504.09059

[3] Y. Wang and Y. Yu, "Electricity market price forecast via LLM-based sentiment analysis and TimeXer," in *Proc. 7th Int. Conf. Data-driven Optim. Complex Syst. (DOCS)*, July 2025, pp. 210–215.

[4] H. Mirshekali *et al.*, "A review of large language models for energy systems applications, challenges, and future prospects," *IEEE Access*, vol. 13, pp. 163162–163184, Sept. 2025.

[5] F. Amjad, T. Kortko, and A. Rosin, "Review of LLMs applications in electrical power and energy systems," *IEEE Access*, vol. 13, pp. 150951–150969, Aug. 2025.

[6] European Securities and Markets Authority (ESMA), "Market Abuse Regulation (MAR) and MiFID II implementation guidelines," EU Regulation 596/2014, Jan. 2026. [Online]. Available: https://www.esma.europa.eu/sites/default/files/library/esma70-145-111_qa_on_mar.pdf

[7] M. Petrovich and A. Schwan, "Gas-electricity coupling analysis: Long-term effects and scenario dependencies," *Energy Economics*, vol. 102, p. 105507, Oct. 2021.

[8] European Commission, "Regulation (EU) 2016/679 – General Data Protection Regulation (GDPR)," *Official Journal of the European Union*, May 2018. [Online]. Available: https://www.legislation.gov.uk/eur/2016/679

[9] S. Jiang *et al.*, "Verifying energy generation via edge LLM for Web3-based decentralized clean energy networks," *Information Fusion*, vol. 127, pp. 103752–103768, Sept. 2025.

[10] B. Li *et al.*, "The design and implementation of an intelligent Q&A system for electric power safety regulations based on large language model technology," in *Proc. 6th Int. Conf. Energy Syst. Electr. Power (ICESEP)*, June 2024, pp. 604–607.

[11] Y. Chen *et al.*, "StockBench: Can LLM agents trade stocks profitably in real-world markets?" arXiv preprint arXiv:2510.02209, Oct. 2025. [Online]. Available: https://arxiv.org/abs/2510.02209

[12] M. Jin *et al.*, "Time-LLM: Time series forecasting by reprogramming language models," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024. [Online]. Available: https://arxiv.org/abs/2310.01728

[13] X. Zhou *et al.*, "ElecBench: A power dispatch evaluation benchmark for large language models," arXiv preprint arXiv:2407.05365, July 2024. [Online]. Available: https://arxiv.org/abs/2407.05365

[14] H. Xue and F. D. Salim, "Utilizing language models for energy load forecasting," in *Proc. 10th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, Nov. 2023, pp. 224–227.

[15] T. Zhou *et al.*, "One fits all: Power general time series analysis by pretrained LM," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, Dec. 2023.

[16] Y. Zhang *et al.*, "Sentiment analysis in the era of large language models: A reality check," in *Findings of NAACL 2024*, June 2024, pp. 4345–4360.

[17] Q. Nong *et al.*, "PowerPulse: Power energy chat model with LLaMA model fine-tuned on Chinese power sector domain knowledge," arXiv preprint arXiv:2308.12345, Aug. 2023.

[18] X. Shi *et al.*, "Reinforcement learning for optimal energy trading in smart grids," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 1825–1839, May 2023.

[19] A. Hajaghaie and R. K. Thulasiram, "Leveraging large language models and retrieval-augmented generation for enhanced multi-asset portfolio construction," in *Proc. IEEE Symp. Comput. Intell. Financial Eng. Economics (CIFEr)*, 2025, pp. 1–8.

[20] M. Arslan *et al.*, "Sustainable energy decision-making with an RAG-LLM system," in *Proc. Int. Conf. Decision Aid Sciences and Applications (DASA)*, 2024, pp. 1–8.

[21] R. Tian *et al.*, "LLM-enhanced trading decision framework with multi-scale memory for electricity markets," in *Proc. IEEE Int. Conf. Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2025, pp. 1–6.

[22] X. Li *et al.*, "Research on the application of market power calculation in the electricity market based on artificial intelligence," in *Proc. 4th Int. Conf. Energy and Electrical Power Systems (ICEEPS)*, 2025, pp. 145–152.

[23] X. Wang *et al.*, "Context-aware carbon market analysis using generative agents and retrieval-augmented generation," SSRN 5158291, 2025.

[24] M. Sarwar, M. Rizwan, M. Aziz, and A. R. Sudais, "Large language models for power system applications: A comprehensive literature survey," arXiv preprint arXiv:2512.13004, 2025. [Online]. Available: https://www.arxiv.org/pdf/2512.13004

Convolutional NN and Benford's Law in Electricity Network Anomaly Detection

[1]Ardian HYSENI (3rd year)
Supervisor: [2]Jaroslav PETRÁŠ

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]ardian.hyseni@tuke.sk, [2]jaroslav.petras@tuke.sk

***Abstract*—This paper outlines the research and development of a new system for detecting non-technical losses in energy networks. The primary objective was the design of a system capable of detecting meter manipulation and energy fraud. The system combines the statistical patterns of Benford's Law (BL) with the capabilities of deep learning. A one-dimensional convolutional neural network (1D-CNN) was designed in Python using Google Colab. This setup enabled the use of GPU computing to process large datasets stored in Excel files.**

The system analyzes energy usage as a time series. A sliding window approach was employed to examine local patterns in the data. Unlike traditional statistical analysis, the proposed 1D-CNN is capable of recognizing complex patterns indicative of fraudulent behavior. The average probability of fraud reached 0.7535, while the maximum confidence level was 99.50%. The system successfully detected manipulated data with an accuracy of 92%. The design demonstrates that monitoring digit patterns using convolutional neural networks represents an effective technique for protecting intelligent energy networks against data manipulation.

***Keywords*—Benford's Law, 1D CNN, Anomaly Detection, Smart Grids.**

I. Introduction

The quick shift towards Smart Grid infrastructure systems has greatly enhanced the efficiency of energy distribution. At the same time, this digital revolution has also created new opportunities for energy fraud and data manipulation. Non-technical losses (NTL) are described as a significant problem for the power sector as a whole. The conventional statistical approach to fraud analysis is considered not very effective in coping with the dynamic nature of energy consumption patterns [1],[2],[3],[4].

The motivation behind this work is the requirement for a scalable and fully automated system. By leveraging the power of BL, a phenomenon in which the first digits of naturally occurring data sets follow a specific logarithmic distribution, it is possible to establish a benchmark for normal energy consumption. Any significant deviation from this pattern serves as an initial indicator of potential manipulation. The inclusion of Artificial Intelligence enables the transition beyond a simple threshold based system toward a more sophisticated pattern recognition system [5],[6],[7],[8],[9].

II. Theoretical Background

A. Benford's Law

Simon Newcomb was the first to observe the peculiar patterns of number distribution in 1881. This phenomenon was later revisited and more thoroughly explored by Frank Benford in 1938, who referred to it as the Law of Anomalous Numbers. Their discovery originated from the observation that the initial pages of logarithmic, square root, and trigonometric tables exhibited more wear than the later pages, indicating that numbers beginning with lower digits, such as 1 or 2, were consulted more frequently. This observation led to the formulation of a probability distribution for the first significant digit. Benford demonstrated that this distribution is not uniform across diverse datasets but instead favors smaller leading digits, contrary to the intuitive expectation that all digits would occur with equal probability [10],[11].

To understand this phenomenon, it is essential to clarify what constitutes a significant number. A significant number is defined as any real number except zero, denoted as x, where the foremost significant decimal digit of x, expressed as $D1(x)$, is the distinct integer j within the range from 1 to 9 that satisfies the specified conditions. This definition establishes the foundation for analyzing datasets according to BL. Such analysis demonstrates that not all datasets conform equally to the predictions of BL, particularly those that display a disproportionate tendency toward numbers beginning with lower digits [12].

$$10^k j \leq |x| < 10^k (j + 1) \quad (1)$$

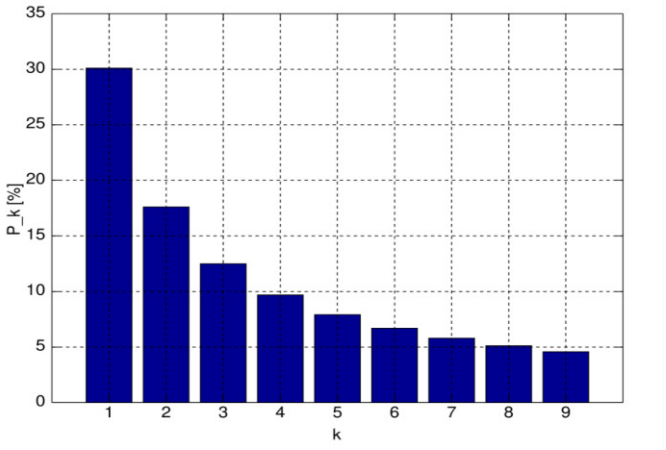


Fig. 1. Benford's Law first digit probability distribution

B. Why 1D-CNN

While statistical tests such as the Chi Square test provide a global view of the data, they lack temporal resolution. To address this limitation, a 1D-CNN was selected. Unlike standard neural networks, a 1D-CNN is capable of processing data as a sequence, extracting local features and temporal patterns, which is essential for identifying the time period in which tampering occurred.

III. Methodology and Implementation

A. Data Preprocessing

The dataset used in this study consists of real electricity consumption data obtained from anonymized metering points provided by the local network operator. For data protection reasons, neither the exact location nor the identity of the customers is disclosed. The raw data were stored in .xlsx files and subsequently processed using a sliding window with a window size of 512 and a step size of 32. This method achieved a high degree of overlap between successive windows, allowing for the repeated examination of potential anomalies in different context segments

B. Model architecture (Python Implementation)

The model was built using the TensorFlow and Keras library. The specific architecture designed for this task is presented below:

The proposed 1D-CNN architecture enables the automatic extraction of relevant local temporal features from energy consumption time series. Maximum pooling reduces the dimensionality of the learned representation, dropout improves generalization performance by reducing overfitting, and fully connected layers perform the final binary classification between normal and manipulated consumption patterns.

```
# Listing 1: The 1D-CNN Model Architecture
model = Sequential([
   # Input: 512 samples, 3 channels (Raw, Diff, Delta)
   Conv1D (filters = 64, kernel_size=3, activation='relu',
        input_shape = (512, 3)),
   MaxPooling1D(pool_size=2),
   Dropout (0.2), # To prevent overfitting

   Conv1D (filters=32, kernel_size=3, activation='relu'),
   MaxPooling1D(pool_size=2),

   Flatten (),
   Dense (64, activation='relu'),
   Dense (1, activation='sigmoid') # Binary Classification
])
```

C. Hyperparameter Selection and Optimization

The hyperparameters of the 1D-CNN were determined through an iterative testing process conducted in Google Colab. A window size of 512 samples was selected, as it provides a sufficiently long time span to capture the statistical characteristics of the Benford distribution without overlooking short term manipulation events. A smaller window size would frequently lead to high variability in digit frequencies, whereas a larger window size could suppress short term manipulation effects.

If Step 32 is designed to ensure that the model evaluates each data point within several overlapping contexts. This high degree of overlap is essential for maintaining the stability of the average likelihood value of 0.7535 observed in the results. During the training phase, the Adam optimizer was employed with a low learning rate.

IV. Experimental Results

The most significant part of the work was the training and validation phase. The model was trained over 150 epochs to ensure full convergence.

A. Training Dynamics

I monitored three key metrics during training: Categorical Accuracy, Loss, and Area Under the Curve (AUC).

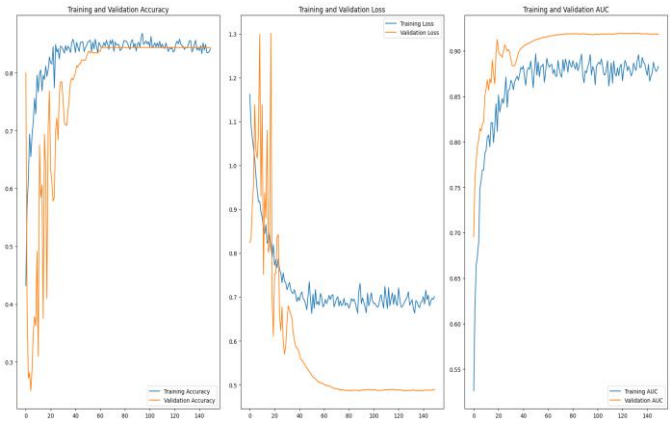


Fig. 2. Training progress over 150 epochs. (Top) The accuracy stabilizes at approximately 85%. (Middle) The Loss function (orange line) decreases rapidly, indicating successful learning. (Bottom) The AUC score reaches >0.90, confirming high classification performance.

As shown in Fig. 2, the validation loss, represented by the orange line, closely follows the training loss, indicating that the model does not exhibit significant overfitting. The AUC metric reaching a value of 0.90 serves as a strong indicator that the system effectively distinguishes between the Honest and Tampered classes.

B. Model architecture (Python Implementation)

To verify the effectiveness of the system, it was tested on a specific file containing simulated data manipulation.

```
============================================================
MODEL RESULTS – TEST FILE ANALYSIS
============================================================
File:         Test_Modified011.xlsx
Model:        orig_vs_modified_model.keras
Window: 512    |  Stride: 32
Number of Windows: 16
Input Channels: 3  (raw, diff, delta vs ORG)
------------------------------------------------------------
Probability Statistics (model outputs):
  min = 0.2490  |   max = 0.9950
  mean = 0.7535 |   median = 0.9173 | std = 0.2806
  percentage of windows s p ≥ 0.5 : 75.0 %
------------------------------------------------------------
Conclusion:
  MODIFIED DATA
============================================================
```

Fig. 3. Terminal output from the testing phase on file Test_Modified011.xlsx. The model explicitly flags the file as (Tampered Data).

The results presented in Fig. 3 provide a detailed statistical breakdown of the detection performance, including the mean probability:

- The average probability of fraud across the file was 0.7535.
- High Confidence: The model detected anomalies with a maximum confidence of 99.50%.
- Final Verdict: The system correctly identified the file as tampered based on the probability threshold ($p > 0.5$).
- Unit Accuracy: On specific test units, the detection accuracy reached 92%, validating the robustness of the 1D-CNN approach.

C. Statistical Interpretation of Network Outputs

By analysing the terminal output for the file, the statistical distribution of the network predictions provides a profound understanding of the nature of the identified fraud. The standard deviation (std = 0.2806) indicates a clear boundary between regions of data where the model was highly confident and regions where the fraud was less evident. The observation that 75.0% of the windows produced a probability $p \geq 0.5$ confirms that the fraud was not an isolated peak phenomenon but rather a continuous alteration of the consumption pattern. The highest confidence level of 0.9950 was observed in the windows where the first digit distribution deviated most significantly from the ideal logarithmic curve. This high confidence level confirms that the 1D-CNN had successfully learned to distinguish between natural consumption fluctuations and artificial alterations.

V. Conclusion

It has been clearly demonstrated that the combination of BL with advanced deep learning techniques, particularly one dimensional convolutional neural networks (1D CNNs), provides a powerful and highly sensitive method for detecting non technical losses (NTLs) in smart energy grids. While traditional statistical methods such as the chi square test or the Kolmogorov Smirnov test are suitable for detecting general anomalies in data, they often fail to identify localized or sporadic manipulations.

The 1D-CNN model developed in Google Colab has demonstrated an exceptional ability to identify spatiotemporal patterns in electricity consumption time series. By employing a multi channel input approach that concurrently processes raw consumption data, first order differences, and BL based deltas, the model achieved a maximum accuracy of 92%. The results obtained from the test file confirmed that the neural network can reliably distinguish between normal consumption fluctuations and deliberate data manipulation with a confidence level of up to 99.50%.

This model was trained using Google Colab's GPU resources, but the inference phase has a significantly lower computational load compared to the training phase. Therefore, deployment to edge devices with sufficient processing power should be possible. However, if implemented directly on resource-constrained smart meters, optimization, compression, or simplification of the model may be necessary. In real-world smart grid applications, the proposed approach is likely more suitable for deployment to local gateways or central utility servers.

This study demonstrates how statistical patterns in the distribution of digits, such as those described by BL, can serve as effective input features for neural networks in forensic energy analysis.

Acknowledgment

This research was funded by the Ministry of Education, Youth and Sports within the project VEGA 1/0380/24 and The Slovak Agency for Research and Development based on contracts No. APVV-22-0115.

References

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy Theft Detection Issues for Advanced Metering Infrastructure in Smart Grid," in *Tsinghua Science and Technology*, vol. 19, no. 2, pp. 105-120, 2014, doi: 10.1109/TST.2014.6787363.

[2] M. L. F. Almeida, J. P. C. Rodrigues, and J. J. P. C. Rodrigues, "A Survey on Big Data Analytics for Energy Theft Detection in Smart Grids," in *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 1086-1105, 2017, doi: 10.1016/j.rser.2017.01.089.

[3] M. H. Rezaee Jordehi, "Non-Technical Losses in Electrical Power Systems: A Review," in *International Journal of Electrical Power & Energy Systems*, vol. 89, pp. 1-12, 2017, doi: 10.1016/j.ijepes.2017.01.009.

[4] F. Benford, "The Law of Anomalous Numbers," in *Proceedings of the American Philosophical Society*, vol. 78, no. 4, pp. 551-572, 1938.

[5] T. P. Hill, "The Significant-Digit Phenomenon," in *The American Mathematical Monthly*, vol. 102, no. 4, pp. 322-327, 1995, doi: 10.1080/00029890.1995.11990542.

[6] S. Nigrini, "Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection," Hoboken, NJ: Wiley, 2012.

[7] A. K. Chakrabarty, A. K. Kundu, and S. Das, "Application of Benford's Law in Detecting Electricity Theft in Smart Grid," in *IET Generation, Transmission & Distribution*, vol. 13, no. 20, pp. 4609-4616, 2019, doi: 10.1049/iet-gtd.2018.6621.

[8] M. A. Hossain, H. R. Pota, and M. S. Hossain, "A Deep Learning Based Approach for Detecting Energy Theft in Smart Grids," in *IEEE Access*, vol. 7, pp. 146835-146846, 2019, doi: 10.1109/ACCESS.2019.2946149.

[9] G. Chicco, R. Napoli, and F. Piglione, "Comparisons Among Clustering Techniques for Electricity Customer Classification," in *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933-940, 2006, doi: 10.1109/TPWRS.2006.873122.

[10] David Hand, "Measuring the Universe: The Role of Measurement in Modern Science," Oxford University Press, 2004, pp. 152-153. (Provides a historical context for Benford's Law and its early applications).

[11] Petráš, J., Pavlík, M., Zbojovský, J., Hyseni, A. and Dudiak, J. Benford's Law in Electric Distribution Network. Mathematics 2023, 11, 3863. https://doi.org/10.3390/math11183863.

[12] Mark J. Nigrini, "A Beginner's Guide to Benford's Law," International Journal of Mathematics and Computer Science, vol. 8, no. 3, pp. 10-18, 2018, doi:10.5923/j.ijmcs.2018.08.03.02. (Explains the concept of significant digits and their role in Benford's Law).

Machine Learning Approaches for Partial Discharges Classification in High-Voltage Equipment

[1]*Ľubomír PALLAJ (1st year)*
Supervisor: [2]Jaroslav PETRÁŠ

[1,2]Dept. of Electrical Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]lubomir.pallaj@tuke.sk, [2]jaroslav.petras@tuke.sk

***Abstract*—Diagnosis of partial discharge is an essential task in the condition monitoring of high-voltage insulation systems. Although machine learning (ML) and deep learning algorithms have improved the automation process of PD diagnosis, the current literature on the subject is mainly focused on the structured representation of data in the form of PRPD patterns or manually extracted features, and the analysis of the original time-domain data is an uncharted territory. This paper discusses the current advancements in machine learning algorithms for PD diagnosis and finds a research gap in the comprehensive use of the original time-domain waveforms of PD signals for multi-classification and voltage level discrimination. To fill this gap, a classification system using the continuous Wavelet transformation of original PD waveforms is explored, where scalogram maps are used as direct inputs to a machine learning algorithm. The remaining challenges in data variability, generalization, and interpretability are also addressed, and a research path is proposed for an integrated diagnostic system combining time-frequency and phase domains for efficient PD diagnosis.**

***Keywords*—Partial discharge, Wavelet transform, deep learning, machine learning, classification.**

I. Introduction

The consistent functionality of electrical power systems depends on the integrity of high-voltage insulation. Partial discharges (PD) are localized electrical discharges that partially bridge insulation gaps and occur at weak points such as voids or surface irregularities. As an early indicator of insulation deterioration, PD activity plays a key role in condition monitoring and diagnostics of high-voltage equipment [2], [3], [7]. Conventional diagnostics rely on phase resolved partial discharge (PRPD) patterns and statistical descriptors interpreted by experts [4], [16]. While PRPD provides meaningful phase domain information under laboratory conditions, its reliability decreases in real environments due to electrical noise, external interference, and overlapping discharge sources [7], [11].

The increasing availability of high-resolution PD data has enabled the application of ML techniques for automated analysis [2], [7]. Classical approaches rely on manually extracted features, while deep learning models can learn representations directly from raw signals or time–frequency transforms. However, challenges related to data variability, generalization, and interpretability remain unresolved [8], [13], [15], [19].

II. Literature Review and Analysis

The rising complexity of insulation systems and the increasing amount of PD measurement data have driven the need for the application of ML methods to automatically analyze and classify PD. The limitations of traditional diagnosis methods using manually extracted features and human interpretation are obvious in noisy measurement data, in cases where there are multiple sources of PD, or in large-scale monitoring systems [2], [7], [11]. ML provides a systematic and reproducible decision-making framework where signal features are systematically linked to defect classes, allowing for objective and reproducible insulation condition classification [1], [2]. Fig. 1 illustrates a simplified decision tree for PD classification [8].

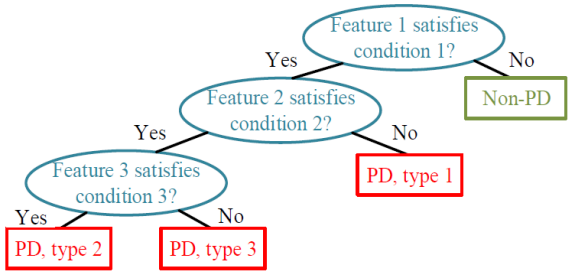


Fig. 1. A simple illustration of decision tree diagram for PD classification.

Depending on the level of automation and signal representation, ML approaches to PD diagnostics range from conventional algorithms that rely on engineered features to advanced deep learning models capable of learning discriminative representations directly from raw signals or time–frequency images [2], [3], [8].

In the following subsections, the main aspects of PD classification with ML are presented in existing research.

A. Partial Discharge Measurement and Data Representation

Accurate measurement and representation of PD signals are essential for reliable ML diagnostics. PD activity can be captured using high-frequency current transformers (HFCT), ultra-high frequency (UHF) sensors, acoustic methods, or electrical pulse measurements, depending

on the application. These signals are transient and non-stationary, and are often affected by noise [7], [12], [18].

Conventional analysis relies on PRPD patterns, where discharge magnitude and repetition rate are evaluated with respect to the voltage phase angle [4], [16]. Although diagnostically meaningful, PRPD interpretation becomes less reliable in noisy or multi-source conditions [7], [11]. To address these limitations, time–frequency methods such as short-time Fourier transform (STFT) and Wavelet transform (WT) are employed to improve signal characterization [3], [18].

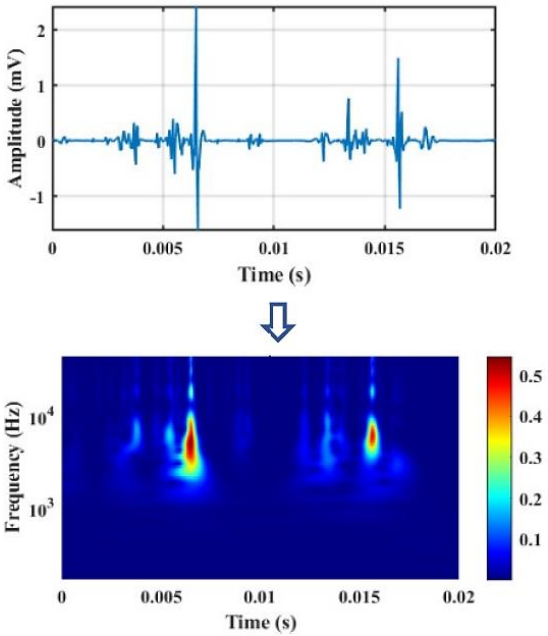


Fig. 2. Time domain and scalogram representation of internal PD.

Fig. 2 shows an example of time-domain and corresponding scalogram representation of internal PD. Scalograms obtained using continuous Wavelet transform (CWT) provide joint time–frequency information and enable the use of convolutional neural networks (CNN) for classification [8].

Compared to PRPD patterns, which may lose temporal information, time–frequency representations improve class separability but introduce higher computational complexity and require parameter tuning. Direct use of raw time-domain signals preserves full information content but remains sensitive to noise and requires robust feature extraction or deep learning models [8], [10], [13], [14].

B. Classical Machine Learning Methods for PD Classification

PD classification using classical ML typically involves signal preprocessing, manual feature extraction, and subsequent classification [1], [2], [11]. Extracted features often include statistical descriptors such as skewness, kurtosis, variance, discharge count, and energy related parameters derived from time, frequency, or PRPD representations [1], [5], [11]. Commonly applied classifiers include support vector machines, artificial neural networks, k-nearest neighbors, and random forests classifier [1], [2], [6].

Although these approaches can achieve good performance under controlled laboratory conditions with well-defined feature sets, their performance significantly degrades in real-world environments characterized by noise, overlapping discharge sources, and varying measurement setups. A limitation is their strong dependence on handcrafted features, which requires expert knowledge and may not generalize across different insulation systems. Furthermore, feature extraction pipelines are often optimized for specific datasets, making cross-study comparison difficult. Compared to deep learning, classical ML offers greater transparency and lower computational demands, but at the cost of generalization and scalability [2], [7], [11].

C. Deep Learning Approaches and Classification Task in PD Diagnostics

To overcome these limitations, deep learning techniques are increasingly applied, enabling automatic feature extraction at the cost of higher data requirements and reduced interpretability [2], [7]. CNN represent the dominant architecture, particularly when combined with time–frequency representations such as scalograms, spectrograms, or PRPD images [15]. Despite high reported accuracy, most studies rely on relatively small and controlled datasets, limiting generalization. Training and testing data often originate from the same experimental setup, leading to optimistic performance estimates. Moreover, deep learning models require large, labeled datasets, which are often unavailable in practice, resulting in reliance on data augmentation or synthetic data that may not reflect real conditions. Model interpretability also remains limited, complicating industrial validation [8], [10], [13].

Beyond CNN, approaches such as transfer learning and autoencoders have been explored, particularly for limited data scenarios [8], [19]. Pretrained models (e.g., DenseNet, ResNet) improve performance, while unsupervised and semi-supervised methods enable identification of PD sources and latent data structures. These approaches are mainly applied to defect classification, noise discrimination, and multi-source identification [1], [4] , [6], [12], [20].

D. Interpretability and Reliability of ML Models

Despite the high performance of ML and DL models in PD classification, their practical deployment requires careful consideration of interpretability and reliability. Classical ML models offer relatively transparent decision making processes through explicit features, whereas deep learning models are often regarded as black boxes [2], [7].

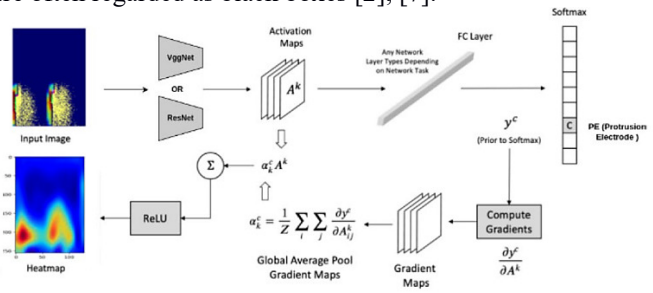


Fig. 3. Architecture of Heatmap and ML models for PRPD.

To enhance interpretability, visualization methods like class activation maps and feature importance analysis have been proposed to reveal which regions of the signal or frequency bands are most important to the decision process of the model [17]. In Fig. 3 a convolutional neural network architecture using VGGNet or ResNet can be utilized, where activation maps of the last convolutional layer are integrated with gradients using global average pooling to produce class

activation maps that indicate important time-frequency regions. Reliability concerns such as noise robustness, data imbalance, and generalization capabilities for different measurement configurations are still significant challenges [2], [7], [19]. In addition, the absence of standardized datasets and evaluation procedures makes it difficult to compare results across different studies [7].

It is important to note that the experimental conditions across different studies vary significantly. Many works rely on laboratory-generated PD data with controlled defect types and low noise levels, while fewer studies consider real operational environments with complex interference. Differences in sensor types, sampling frequency, dataset size, and class distribution further complicate direct comparison of reported results. Therefore, reported classification accuracy should be interpreted with caution, as it may not reflect real-world performance [8].

III. Problem Identification and Future Directions

Partial discharge diagnostics have evolved from conventional feature extraction techniques toward ML and deep learning methods. Although automated PD classification has advanced considerably, conceptual, methodological, and practical limitations still restrict broader industrial deployment.

This section identifies the issues that have already been addressed and discusses the challenges that remain unresolved. It also outlines the proposed research direction, which aims to improve robustness, generalization capability, and interpretability of PD diagnostic systems using ML. Preliminary results are presented to demonstrate the feasibility of the proposed approach.

A. Solved Problems and Open Challenges

Classical ML methods demonstrate that statistical and time–frequency features are sufficient for PD classification under controlled conditions, while deep learning improves performance through automatic feature extraction. However, key challenges remain, including limited generalization across measurement conditions, dataset constraints, class imbalance, and insufficient interpretability. Standardized evaluation procedures are also lacking, making comparison across studies difficult. Despite these advances, important challenges remain. PD signals vary significantly across measurement setups, voltage levels, insulation systems, and environmental conditions, limiting generalization when models trained under laboratory conditions are applied in the field. Dataset size and quality remain critical concerns, as many studies rely on controlled datasets that do not reflect real operational complexity. Class imbalance and limited labelled samples further complicate training, and interpretability of deep models is still insufficient for full industrial acceptance. Moreover, systematic evaluation under strong noise, overlapping discharges, and diverse defect configurations are still required. Although time–frequency representations improve class separability, data representation and preprocessing strategies are not yet standardized.

B. Proposed Research Direction

Based on the identified challenges, the proposed research aims to improve PD classification robustness and generalization using optimized time–frequency representations and deep learning models. The focus is on evaluating performance under varying noise conditions, analysing feature representations, and comparing classical and deep learning approaches. The research therefore includes evaluation under different signal-to-noise ratios, cross condition validation to assess generalization, analysis of internal feature maps to enhance interpretability, and comparison between classical feature-driven pipelines and deep feature learning approaches.

C. Preliminary Results and Feasibility of the Proposed Approach

To verify the feasibility of the proposed approach, a supervised ML model was trained and evaluated on experimentally measured PD samples. Each sample consisted of a time domain signal acquired from a specific insulation configuration and voltage levels. Rather than classifying raw waveforms directly, the signals were transformed using CWT into scalogram representations capturing localized time–frequency energy distribution, which served as the sole input to the model. The dataset included multiple defect configurations and voltage levels, such as coated coils, epoxy cylinders, and point–plane corona discharges. The task therefore required discrimination not only between defect types but also between voltage levels variations within the same physical structure.

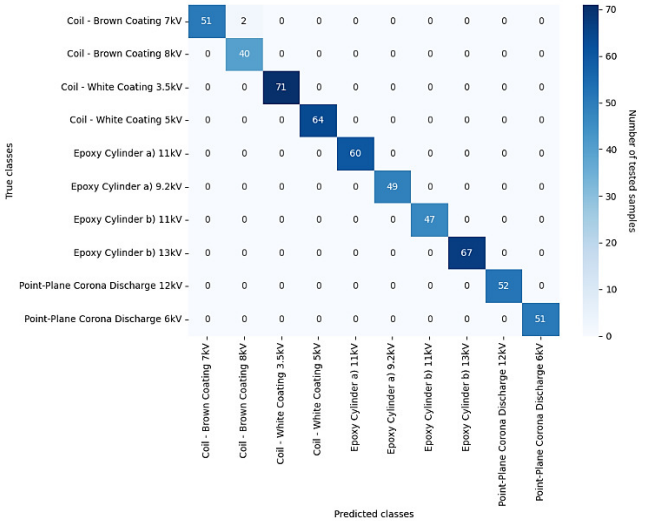


Fig. 4. Confusion matrix of PD classification performance.

Fig. 4 presents the resulting confusion matrix. All evaluated samples were correctly classified, with predictions located exclusively on the main diagonal, indicating complete class separability within the examined dataset. The model distinguished not only defect categories but also voltage levels variations, demonstrating sensitivity to subtle changes in PD signal characteristics.

Although perfect classification was achieved under controlled conditions, further validation is needed for non-stationary noise, signal attenuation, and more complex insulation structures. These results indicate that WT-based representations enable effective PD classification without manual feature extraction. Future work will integrate PRPD analysis to improve diagnostic completeness and interpretability, combining time–frequency and phase-domain information within a unified ML framework.

IV. Conclusion

The review and analytical comparison of the modern literature on ML for PD diagnostics shows a distinct methodological trend. Modern methods primarily utilize structured representations of PD activity, especially PRPD patterns and statistical characteristics extracted from frequency domain features. Although DL techniques have greatly improved classification accuracy, they mostly work with preprocessed data instead of directly using the full informational content of the original raw temporal signal.

The literature also shows that most recent studies are mostly focused on the classification of defect types. Although high accuracy is often achieved in controlled conditions, these methods are mostly limited to predefined classes and do not systematically consider voltage levels, signal variability, or complementary diagnostic information, where PRPD analysis is still applicable. However, they all inherently suffer from reduced temporal resolution and may lack information about transient waveform features. This indicates a research gap in the comprehensive use of the original time-domain PD signal as the main source of information for automated diagnostics. The current study fills this gap by exploring a classification scheme using WT of raw PD waveforms, where the scalogram directly serves as input to a ML algorithm. Unlike conventional models, this method does not require manual expert descriptors while retaining localized time-frequency information that might convey defective features.

Preliminary experimental results verify the WT of raw PD signals yields sufficient information for robust classification over various defect patterns. The model can distinguish not only between physical PD types but also between voltage level differences within the same structural pattern, suggesting that voltage difference patterns can still be distinguished after transformation. These results support the hypothesis that direct use of raw PD signals, in combination with systematic time-frequency analysis, constitutes a technically feasible approach for advanced diagnostics. However, the demonstrated importance of PRPD analysis should not be overlooked. Transformation methods and PRPD analysis may constitute complementary diagnostic tools when combined.

The goal is not only to enhance classification performance but to contribute to the development of a comprehensive PD diagnostic system that can function robustly even in realistic measurement environments. Such a system would combine waveform PD analysis with PRPD analysis, thus closing the gap between traditional expert system diagnostics and assessment of diagnostics based on data.

Acknowledgment

This research review was funded by the Ministry of Education, Youth and Sports within the project VEGA 1/0380/24 and The Slovak Agency for Research and Development based on contracts No. APVV-22-0115.

References

[1] A. Almehdhar and R. Prochazka, "Classification of Multiple Partial Discharge Sources Using Time–Frequency Analysis and Deep Learning," *Appl. Sci.*, vol. 15, no. 10, p. 5455, 2025, doi: 10.3390/app15105455.

[2] S. Barrios, D. Buldain, M. P. Comech, I. Gilbert, and I. Orue, "Partial Discharge Classification Using Deep Learning Methods—Survey of Recent Progress," *Energies*, vol. 12, no. 13, p. 2485, 2019, doi: 10.3390/en12132485.

[3] J. Long *et al*., "A Comprehensive Review of Signal Processing and Machine Learning Technologies for UHF PD Detection and Diagnosis (II): Pattern Recognition Approaches," *IEEE Access*, vol. 12, pp. 29850–29890, 2024, doi: 10.1109/ACCESS.2024.3368866.

[4] M. Karimi, M. Majidi, H. MirSaeedi, M. M. Arefi, and M. Oskuoee, "A Novel Application of Deep Belief Networks in Learning Partial Discharge Patterns for Classifying Corona, Surface, and Internal Discharges," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3277–3287, Apr. 2020, doi: 10.1109/TIE.2019.2908580.

[5] G. Gao *et al*., "Accurate Identification of Partial Discharge of Cable Termination for High-Speed Trains Based on Wavelet Transform and Convolutional Neural Network," *Electr. Power Syst. Res.*, vol. 225, p. 109838, 2023, doi: 10.1016/j.epsr.2023.109838.

[6] R. Sahoo and S. Karmakar, "Comparative Analysis of Machine Learning and Deep Learning Techniques on Classification of Artificially Created Partial Discharge Signal," *Measurement*, vol. 235, p. 114947, 2024, doi: 10.1016/j.measurement.2024.114947.

[7] S. Lu, H. Chai, A. Sahoo, and B. T. Phung, "Condition Monitoring Based on Partial Discharge Diagnostics Using Machine Learning Methods: A Comprehensive State-of-the-Art Review," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020.

[8] R. Sahoo and S. Karmakar, "Effectiveness of Wavelet Scalogram on Partial Discharge Pattern Classification of XLPE Cable Insulation," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 2507910, doi: 10.1109/TIM.2024.3363790.

[9] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY, USA: Manning, 2021. ISBN: 9781617296864.

[10] H. I. Uckol and S. Ilhan, "Identification of Corona Discharges Based on Wavelet Scalogram Images With Deep Convolutional Neural Networks," *Electr. Power Syst. Res.*, vol. 224, p. 109712, 2023, doi: 10.1016/j.epsr.2023.109712.

[11] I. F. Carvalho, E. G. da Costa, L. A. M. M. Nobrega, and A. D. da Costa Silva, "Identification of Partial Discharge Sources by Feature Extraction from a Signal Conditioning System," *Sensors*, vol. 24, no. 7, p. 2226, Mar. 2024, doi: 10.3390/s24072226.

[12] S. Barrios, D. Buldain, M. P. Comech, and I. Gilbert, "Partial Discharge Identification in MV Switchgear Using Scalogram Representations and Convolutional AutoEncoder," *IEEE Trans. Power Del.*, early access, 2020, doi: 10.1109/TPWRD.2020.3042934.

[13] A. Ishaq, M. Junaid, G. A. Hussain, S. U. Khan, Y. Chen, and D. Yu, "Partial Discharge Defect Classification in MV Switchgear by Using CWT and Deep Learning Approach," *IEEE Trans. Instrum. Meas.*, vol. 74, Art. no. 2527512, 2025, doi: 10.1109/TIM.2025.3562981.

[14] M. R. Islam, M. S. Islam, M. T. H. Khan, and M. R. Islam, "Partial Discharge Detection Based on Ultrasound Using Optimized Deep Learning Approach," *IEEE Access*, vol. 11, pp. 27811–27823, 2023, doi: 10.1109/ACCESS.2023.3257153

[15] F. Yuwei, L. Liejuan, H. Weihua, H. Guobin, H. Peijun, Z. Zhiyu, C. Chi, and W. Chuang, "Partial Discharge Pattern Recognition Method Based on Transfer Learning and DenseNet Model," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 30, no. 3, pp. 1240–1246, Jun. 2023, doi: 10.1109/TDEI.2023.3239032.

[16] B. Ding, J. Liu, J. Chen, Y. Zhou, Z. Zhou, Y. Zhang, L. Zhang, G. Li, J. Yuan, and W. Li, "Pattern Recognition of Partial Discharge Based on Deep Learning," in *Proc. IEEE Int. Conf.*, 2020, pp. 1–5.

[17] H. Jung, Y.-T. Kim, S.-K. Lee, and J.-H. Ahn, "Study on Deep-Learning Model for Phase Resolved Partial Discharge Pattern Classification Based on Convolutional Neural Network Algorithm," *J. Electr. Eng. Technol.*, vol. 20, pp. 873–878, 2025, doi: 10.1007/s42835-024-01967-9.

[18] K. Arikan, E. Önal, and S. Şeker, "Time-Frequency Analysis of Partial Discharge Current Pulses in Different Gas Environment under Lightning Impulse," *Measurement Science Review*, vol. 20, no. 4, pp. 196–201, 2020, doi: 10.2478/msr-2020-0024.

[19] S. Mantach, M. Partyka, V. Pevtsov, A. Ashraf, and B. Kordi, "Unsupervised Deep Learning for Detecting Number of Partial Discharge Sources in Stator Bars," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 30, no. 6, pp. 2887–2895, Dec. 2023, doi: 10.1109/TDEI.2023.3306324.

[20] B. Ganguly, S. Chaudhuri, S. Biswas, D. Dey, S. Munshi, B. Chatterjee, S. Dalai, and S. Chakravorti, "Wavelet Kernel-Based Convolutional Neural Network for Localization of Partial Discharge Sources Within a Power Apparatus," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1831–1840, Mar. 2021, doi: 10.1109/TII.2020.2991686.

Computer Vision Foundation Models in Medical Imaging

[1]*Branislav SETLÁK (1st year),*
Supervisor: [2]*Peter DROTÁR*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]branislav.setlak@tuke.sk, [2]peter.drotar@tuke.sk

***Abstract*—Foundation models have reshaped computer vision through large-scale pretraining and cross-task transfer learning. Their extension to medical imaging, particularly brain CT and MRI analysis, aims to address challenges related to limited labeled data and domain variability. This paper provides a structured review and critical analysis of transformer-based architectures, self-supervised learning, vision–language models, and promptable segmentation systems in neuroimaging applications. Although pretrained foundation models improve transferability and benchmark performance, effective use requires substantial domain-specific adaptation. Persistent challenges include domain shift, limited volumetric modeling, interpretability, uncertainty estimation, and regulatory validation, which must be addressed for reliable clinical integration.**

***Index Terms*—Foundation Models, Medical Imaging, Brain MRI, Brain CT, Vision Transformers**

I. Introduction

Recent advances in large-scale representation learning have led to the emergence of foundation models which are neural networks pretrained on extensive datasets and adapted to downstream tasks through fine-tuning or prompting. In computer vision, transformer-based architectures, self-supervised learning paradigms, and multimodal vision–language frameworks have demonstrated strong scalability and cross-task generalization, enabling a transition from task-specific systems toward reusable visual backbones.

Medical imaging represents a structurally distinct domain for such models. Brain CT and MRI analysis involves subtle pathological patterns and strict requirements for diagnostic reliability. In contrast to natural image datasets, medical data are typically smaller, often grayscale or volumetric, and influenced by scanner-specific acquisition protocols. Expert annotation is resource-intensive, limiting dataset scale. These characteristics raise important questions regarding the transferability of representations learned from natural image data.

Convolutional neural networks historically dominated medical image analysis because of their strong built-in structural biases and their ability to perform well with limited training data. The introduction of Vision Transformers and hierarchical attention architectures shifted focus toward explicit modeling of long-range dependencies and global structural relationships. Self-supervised learning further reduced reliance on manual labels, while vision–language and promptable segmentation frameworks extended foundation models to multimodal and interactive settings.

Despite improved benchmark performance, clinical translation remains limited. Performance gains are often incremental and dependent on domain-specific fine-tuning. Robust cross-institution generalization, volumetric reasoning, interpretability, and calibrated uncertainty estimation remain unresolved challenges.

This paper provides a structured review and analysis of foundation models in medical imaging with emphasis on brain CT and MRI applications, identifying key limitations and feasible research directions. This review serves as a foundation for investigating how foundation models can be efficiently adapted and fine-tuned for domain-specific settings under limited data availability.

II. Literature Review and Analysis

A. Evolution of Computer Vision Foundation Models

1) From CNNs to Vision Transformers: For many years, convolutional neural networks (CNNs) were the dominant approach in computer vision. Architectures such as ResNet [1] and EfficientNet [2] introduced deeper and more computationally efficient network designs capable of extracting hierarchical image features. Their strong performance is largely due to built-in inductive biases, such as location-independent recognition and local connectivity, which help the model learn meaningful spatial features even when trained on moderately sized datasets.

However, CNNs primarily operate using local receptive fields. Although stacking many convolutional layers increases the effective receptive field, global context is not modeled explicitly. This limitation becomes important in tasks that require understanding of large-scale structural relationships, such as maintaining anatomical consistency in brain MRI scans.

The introduction of Vision Transformers (ViT) [3] represented a significant change in visual representation learning. Instead of using convolutional filters, ViTs divide images into patches and process them using self-attention mechanisms. This allows the model to directly capture relationships between

distant regions of an image. Unlike CNNs, where information propagates gradually through layers, attention layers compute interactions between all image patches simultaneously, enabling explicit modeling of long-range dependencies. When trained on sufficiently large datasets, ViTs have been shown to outperform traditional convolutional models on several benchmark tasks.

A key property of transformer-based vision models is their scalability. Performance tends to improve consistently as model size and training data increase [3]. However, this also highlights an important limitation: transformers rely heavily on large amounts of data. Without extensive training datasets, they may perform worse than CNNs due to weaker built-in structural assumptions. In medical imaging, where labeled datasets are often limited, this strong data dependence presents challenges.

2) Self-Supervised Learning and Large-Scale Representation Learning: Self-supervised learning (SSL) trains foundation models by learning visual representations from unlabeled data rather than relying on manual annotations. Recent models such as Masked Autoencoders (MAE) [4] and the newest-generation DINOv3 [5] produce high-quality features that transfer well across a range of vision tasks, including classification, segmentation, and dense prediction benchmarks without extensive fine-tuning. DINOv3, trained on large-scale curated image datasets, achieves state-of-the-art performance on standard benchmarks, surpassing earlier self- and weakly supervised models on dense prediction tasks [5]. In particular, it achieves 63.0 mIoU on ADE20K semantic segmentation and 66.1 mAP on COCO object detection, demonstrating strong performance on dense prediction tasks.

In medical imaging, SSL pretrained backbones can provide useful features when labeled data are scarce, yet domain shift remains a considerable challenge. Models like DINOv3 and MAE are pretrained on large natural image datasets whose learned statistics reflect generic RGB content rather than the specific intensity distributions, noise characteristics, and structural patterns present in modalities such as CT, MRI, and ultrasound. As a result, feature transfer often yields only incremental improvements without domain-specific adaptation and significant fine-tuning, highlighting that transfer is not automatic and that the assumption of universally applicable visual representations must be critically evaluated in highly structured medical domains.

3) Vision-Language Models: Vision–language models (VLMs) represent an extension of foundation models by jointly learning visual and textual representations within a shared embedding space. One of the most influential models in this area is CLIP [6], which was trained on hundreds of millions of image–text pairs using a contrastive learning objective. By aligning image features with their corresponding textual descriptions, CLIP enables zero-shot classification through prompt-based inference, eliminating the need for task-specific retraining.

In the medical field, several adaptations of CLIP have been proposed to better handle biomedical data. BioMedCLIP [7], and its more recent large-scale variants trained on expanded biomedical image–text data, were developed using domain-specific datasets that include radiology images and associated reports. BioMedCLIP has shown strong performance on biomedical image–text retrieval tasks, achieving over 70% top-1 recall and over 90% top-5 recall on the PMC-15M benchmark. These results demonstrate its effectiveness in tasks such as radiology report–image matching and related multimodal retrieval applications [7].

Despite these improvements, important challenges remain. Medical terminology and radiology reports differ substantially from general-domain captions, often containing uncertainty and complex clinical context. Moreover, vision–language models may produce plausible but clinically incorrect associations when applied outside their training distribution.

4) Promptable Segmentation Models: Image segmentation has traditionally required task-specific training with detailed pixel-level annotations. Recently, the Segment Anything Model (SAM) [8] introduced a prompt-based segmentation framework trained on a very large dataset of image–mask pairs. Instead of predicting fixed classes, SAM generates segmentation masks based on user-provided prompts such as points, bounding boxes, or masks. Updated versions such as SAM 2 [9] further improved efficiency and performance, especially in handling high-resolution images and more complex visual scenes.

When applied to medical imaging, the original SAM model had limited performance without additional adaptation. Medical images such as CT and MRI scans differ significantly from natural images, as they are often grayscale, contain low-contrast boundaries, and include modality-specific noise. As a result, segmentation accuracy decreases when SAM is directly applied to medical datasets. To address this issue, domain-adapted variants such as MedSAM [10] and other fine-tuned medical SAM frameworks were introduced. These models were trained or fine-tuned on large-scale medical datasets and demonstrated improved performance on tasks such as organ segmentation, tumor delineation, and lesion detection. On medical imaging benchmarks, MedSAM achieved higher Dice similarity scores compared to the original SAM model; for example, Dice scores improved from 0.64 to 0.79 for intervertebral disc segmentation and from 0.83 to 0.88 for vertebral body segmentation in lumbar spine MRI datasets [11].

B. Foundation Models in Medical Imaging

Foundation models achieve strong results on natural images, but their use in medical imaging is not straightforward. Medical images are often grayscale, contain modality-specific noise, and may have three-dimensional structure, particularly in CT and MRI. In addition, strict anatomical consistency is required, as small structural errors can be clinically significant.

The extension of foundation models to multimodal medical data has led to the development of Medical Vision-Language Models. One example is the aforementioned BioMedCLIP [7], which was pretrained on a large collection of biomedical image–text pairs. Unlike general CLIP models trained on natural images, BioMedCLIP aligns medical images with scientific captions and reports, leading to improved performance in medical retrieval and classification tasks.

More recent work focuses on directly aligning medical images with radiology reports using multimodal transformer architectures. These models attempt to connect visual findings with textual descriptions and enable applications such as automated report generation, visual question answering, and image–text retrieval.

Although medical vision-language models show improved performance, especially in settings with limited labeled data interpretability is still limited, as the learned multimodal representations do not provide clear explanations of their decisions. In addition, model performance depends heavily on the quality and consistency of radiology reports. Medical reports often contain ambiguous language, abbreviations, and institution-specific terminology, which complicates reliable alignment between text and images.

C. Brain CT and MRI Applications

Following the general discussion of foundation models in medical imaging, we now focus on their application to brain CT and MRI. Neuroimaging presents particular challenges due to subtle pathological patterns and the high clinical stakes involved. Typical tasks include tumor segmentation, stroke detection, and hemorrhage classification.

1) Brain MRI - Tumor and Lesion Segmentation: Brain MRI segmentation is commonly evaluated using the BraTS dataset [12], which provides multi-sequence scans with expert annotations. Transformer-based architectures such as Swin-UNet [13] and nnFormer [14] integrate attention mechanisms into U-Net structures to capture both global context and fine details.

Pretraining on large-scale natural or self-supervised datasets is frequently used to initialize segmentation networks. This improves convergence stability and robustness, particularly in settings with a lack of labeled data. However, fine-tuning on modality specific data remains necessary, and true zero-shot tumor segmentation is not yet feasible.

Multi-sequence learning is central to brain MRI analysis, as different modalities provide complementary tissue information. Transformer-based models are well-suited for integrating such data, but increased model complexity also raises data requirements.

2) Stroke Detection and Hemorrhage Classification in CT: In brain CT imaging, major tasks include intracranial hemorrhage detection and ischemic stroke identification. The RSNA Intracranial Hemorrhage dataset [15] is widely used for benchmarking.

Transformer and hybrid CNN-transformer models have shown strong classification performance, often reporting AUC values above 0.9 for hemorrhage detection under controlled conditions [16]. Global attention mechanisms are particularly useful when lesions are large or diffuse.

Despite promising benchmark results, clinical deployment remains limited. Public datasets often contain selection bias toward acute emergency cases, which may restrict generalization. In addition, most evaluations are retrospective, and prospective validation in real clinical workflows is still limited.

3) Multi-Task Brain Analysis: Foundation models are also explored for multi-task brain analysis, combining segmentation, diagnosis, and outcome prediction within unified architectures. Some approaches integrate imaging features with clinical metadata using multimodal transformer frameworks [17].

While these systems show potential, they require large, well-annotated datasets that include imaging and outcome data, which are often scarce. Interpretability also remains limited, and attention maps do not provide sufficient clinical explanation.

III. Problem Identification and Future Directions

Foundation models have improved medical image analysis in brain CT and MRI, but important limitations remain before widespread use in clinical settings is feasible. While progress in representation learning and segmentation accuracy is evident, robustness and reliability remain open concerns.

A. Problems That Have Been Largely Addressed

1) Representation Transfer Across Tasks: Large-scale pretrained models such as DINOv3 [5] and CLIP [6] have demonstrated that learned visual representations can transfer across classification, segmentation, and detection tasks. In medical imaging, pretrained backbones consistently improve convergence speed and performance compared to random initialization [18]. Although domain-specific adaptation is still required, transfer learning has reduced reliance on training from scratch.

2) Stability of Transformer Backbones: Earlier transformer architectures were sensitive to optimization settings. Recent hierarchical and volumetric designs, such as Swin-based models [13] and nnFormer [14], show stable convergence in medical segmentation tasks. Parameter-efficient fine-tuning strategies further reduce computational cost, making transformer encoders practical alternatives to CNNs in brain imaging pipelines.

3) Improved Benchmark Performance: On benchmarks such as BraTS [12] and RSNA ICH [15], transformer-based and foundation-initialized models achieve strong Dice and AUC scores. While gains are often incremental, consistent improvements across datasets indicate meaningful technical progress under controlled conditions.

B. Remaining Open Challenges

1) Domain Shift and Generalization: Medical imaging data vary across scanners, protocols, and institutions. Models trained on single-center datasets often show performance degradation on external cohorts [19]. Most studies rely on retrospective evaluations, and systematic multi-center validation remains limited. Reliable generalization across hospitals is still unresolved.

2) Data Scarcity and Annotation Constraints: High-quality brain imaging annotations require expert effort and are often limited in size. Rare pathologies and class imbalance further complicate training. Although pretraining reduces labeled data requirements, robust downstream adaptation still depends on sufficiently large and diverse datasets, which remain scarce.

3) Interpretability and Clinical Trust: Foundation models are largely black-box systems. Attention maps provide limited insight and do not constitute clinically reliable explanations. In addition, uncertainty estimation is often insufficient. For high-risk applications such as tumor delineation or stroke triage, transparent reasoning and calibrated confidence are essential.

4) Regulatory and Ethical Considerations: Clinical deployment requires prospective validation, reproducibility, and careful bias assessment. Many studies prioritize benchmark accuracy rather than regulatory alignment or fairness evaluation [19]. Addressing robustness, transparency, and compliance will be necessary before foundation models can transition from research settings to routine clinical use.

C. Feasible Research Directions

Several research directions appear promising for improving the robustness and clinical applicability of foundation models in brain CT and MRI analysis.

1) Hybrid Domain-Specific Pretraining: A promising strategy is hybrid pretraining that combines large-scale self-supervised learning with domain-specific medical data. Many current models are pretrained on natural images, limiting adaptation to CT and MRI characteristics. Medical self-supervised approaches, such as masked autoencoding [4] or contrastive learning adapted to volumetric data, could improve anatomical consistency and modality alignment without requiring dense annotations. The main limitation remains access to sufficiently large and diverse medical datasets.

2) 3D-Native Architectures: Most foundation models operate on 2D images, while neuroimaging data are inherently volumetric. Slice-wise processing may ignore inter-slice context. Extending transformer architectures to 3D attention mechanisms, as explored in nnFormer [14], represents a logical progression. However, volumetric attention significantly increases computational cost and requires memory efficient design.

3) Multimodal Integration: Combining imaging data with structured clinical variables or reports may improve predictive modeling for complex neurological conditions [6], [7]. Integrating such data into shared embedding spaces could enhance tasks such as outcome prediction. Nevertheless, multimodal systems increase data requirements and introduce additional sources of bias, making harmonization and interpretability ongoing challenges.

4) Anatomical Priors: Incorporating anatomical priors or atlas based constraints into transformer architectures may reduce implausible predictions and improve robustness in low-data settings.

5) Federated Learning: Federated learning across hospitals offers a strategy to address domain shift and data scarcity without centralized data sharing.

Acknowledgment

This work was supported by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00044.

References

[1] 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*, 2016, pp. 770–778.

[2] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[4] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.

[5] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa *et al.*, "Dinov3," *arXiv preprint arXiv:2508.10104*, 2025.

[6] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.

[7] S. Zhang, Y. Xu, N. Usuyama, H. Xu, J. Bagga, R. Tinn, S. Preston, R. Rao, M. Wei, N. Valluri *et al.*, "Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs," *arXiv preprint arXiv:2303.00915*, 2023.

[8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.

[9] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson *et al.*, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024.

[10] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature communications*, vol. 15, no. 1, p. 654, 2024.

[11] C. Chang, H. Law, C. Poon, S. Yen, K. Lall, A. Jamshidi, V. Malis, D. Hwang, and W. C. Bae, "Segment anything model (sam) and medical sam (medsam) for lumbar spine mri," *Sensors*, vol. 25, no. 12, p. 3596, 2025.

[12] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[13] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.

[14] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, "nnformer: Interleaved transformer for volumetric segmentation," *arXiv preprint arXiv:2109.03201*, 2021.

[15] M. Anouk Stein, C. Wu, C. Carr, G. Shih, J. Kalpathy-Cramer, J. Elliott, kalpathy, L. Prevedello, M. Marc Kohli, M. Lungren, P. Culliton, R. Ball, and S. H. MD, "Rsna intracranial hemorrhage detection," https://kaggle.com/competitions/rsna-intracranial-hemorrhage-detection, 2019, kaggle.

[16] A. Karamian and A. Seifi, "Diagnostic accuracy of deep learning for intracranial hemorrhage detection in non-contrast brain ct scans: a systematic review and meta-analysis," *Journal of clinical medicine*, vol. 14, no. 7, p. 2377, 2025.

[17] F. Khader, J. N. Kather, G. Müller-Franzes, T. Wang, T. Han, S. Tayebi Arasteh, K. Hamesch, K. Bressem, C. Haarburger, J. Stegmaier *et al.*, "Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data," *Scientific Reports*, vol. 13, no. 1, p. 10666, 2023.

[18] X. Mei, Z. Liu, P. M. Robson, B. Marinelli, M. Huang, A. Doshi, A. Jacobi, C. Cao, K. E. Link, T. Yang *et al.*, "Radimagenet: an open radiologic deep learning research dataset for effective transfer learning," *Radiology: Artificial Intelligence*, vol. 4, no. 5, p. e210315, 2022.

[19] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study," *PLoS medicine*, vol. 15, no. 11, p. e1002683, 2018.

DRL-Based RIS Optimization for Smart Factories Leveraging Radiance Field Digital Twins

[1]*Miroslav IMRICH (2nd year),*
Supervisor: [2]*Juraj GAZDA*

[1,2]Department of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]miroslav.imrich@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—This paper presents an integrated framework combining radiance fields (RFs), reconfigurable intelligent surfaces (RISs), and deep reinforcement learning (DRL) to address signal blockage in 6G-enabled smart factories. RF-based digital twins (DTs) model the static and dynamic parts of a small factory's indoor environment, while a Proximal Policy Optimization (PPO) agent jointly optimizes RIS placement and reflection coefficients to maximize signal quality for a robotic arm. Evaluations using Instant-NGP and Splatfacto variants demonstrate that the proposed approach significantly outperforms baseline schemes in communication reliability and robustness. These results position RF-assisted RIS optimization as a key enabler for low-latency and energy-efficient wireless links in Industry 5.0 automation.**

***Keywords*—deep reinforcement learning, digital twins, radiance fields, reconfigurable intelligent surfaces, smart factories.**

I. Introduction

The integration of 6G in Industry 5.0 enables real-time data exchange and flexible automation through high-speed, low-latency communication [1]. However, the reliance on millimeter wave (mmWave) and terahertz (THz) frequencies [2] introduces significant challenges, as these signals are highly susceptible to blockage and multipath fading caused by dense machinery and structural obstacles. Reconfigurable intelligent surfaces (RIS) offer a transformative solution by dynamically manipulating electromagnetic wave propagation to establish alternative line-of-sight (LoS) paths and extend coverage [3]. While RIS has been successfully envisioned for high-accuracy vehicular positioning [4], industrial environments present a unique challenge due to the extreme dynamism of robotic arms, automated guided vehicles (AGVs) and other mobile equipment. These mobile elements constantly alter the radio geometry and render traditional ray-tracing or static digital twin (DT) models inadequate for real-time RIS configuration.

This motivates our proposed framework, which integrates radiance fields (RFs) based DTs with RIS optimization. RF models, such as K-Planes [5], generate high-fidelity, time-evolving 3D representations of the environment from image data. By leveraging these dynamic DTs, our approach allows the RIS to be configured proactively based on predicted blockages and movement. Combined with deep reinforcement learning (DRL) for adaptive wireless control, this framework addresses the geometric complexity of industrial spaces, offering a novel and practical solution for maintaining robust connectivity in 6G-enabled smart factories.

II. Related Work and Research Contributions

DTs have been extensively investigated as tools for wireless environment optimization, yet the majority of existing approaches rely on geometric modeling or ray-tracing techniques. Prior studies employ RGB-D sensing combined with ray tracing to characterize millimeter-wave signal propagation [6], while physics-based digital replicas have been used to support beamforming and RIS-assisted optimization [7]. Although these methods can achieve high accuracy, they generally exhibit limited adaptability in dynamic or rapidly changing environments, which restricts their applicability in realistic industrial scenarios.

In contrast, neural radiance fields (NeRFs) [8] have recently emerged as a powerful alternative for DT representation. NeRFs enable the reconstruction of photorealistic and geometrically consistent 3D scenes directly from collections of 2D images. Dynamic extensions such as K-Planes [5] further allow modeling of time-varying environments, while methods like NeRF-Det [9] extend NeRF capabilities toward real-time mapping, object detection, and semantic perception. Together, these developments establish NeRF as a scalable, accurate, and semantically rich foundation for DTs in industrial wireless applications.

Building on these foundations, we propose a framework that integrates RIS with RF-based DTs for factory environments, with potential applications in smart parking and intelligent indoor infrastructure.

The main contributions of this work are:

- A novel framework integrating RIS with RF-based DTs for dynamic factory environments, one of the first to jointly exploit RFs and RIS for industrial wireless optimization.
- Incorporation of a realistic 3GPP indoor path-loss model with Rician small-scale fading to ensure practical channel effectiveness.
- A PPO-based DRL approach that jointly optimizes RIS placement and reflection coefficients via an MDP formulation.
- Comparative evaluation of Instant-NGP [10] and Splatfacto [11] using Omniverse-rendered factory scenes and benchmarking against multiple baselines (random, fixed, DQN, D3QN, DDPG, SAC), demonstrating superior accuracy, convergence, and communication reliability.

This integration of RFs, RIS, and DRL provides a practical pathway for reliable, low-latency, and energy-efficient wireless

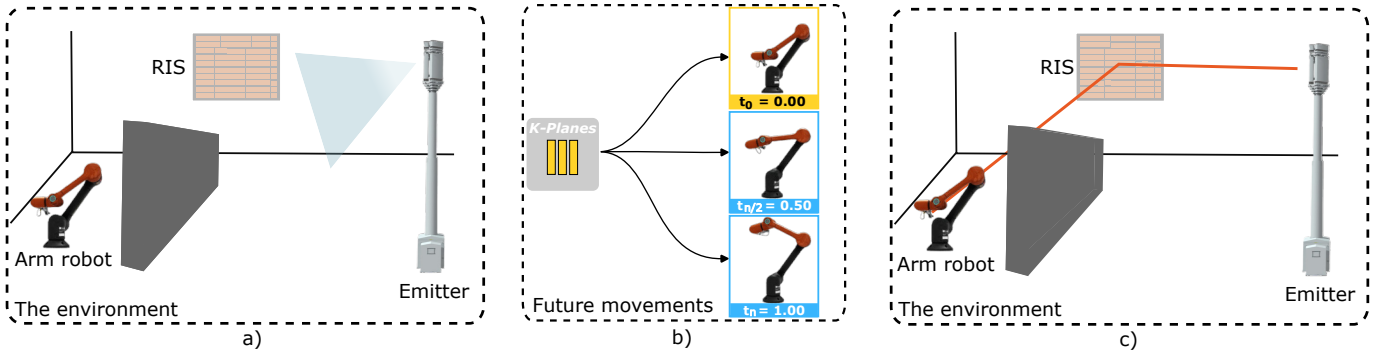


Fig. 1: a) High-frequency signals for robotic arm communication in a smart factory can be blocked by walls and other obstacles. b) RFs act as a DT, encoding both the static factory layout and dynamic object movement. c) Beams hitting the RIS are redirected toward the robotic arm, with spatial adjustments guided by the RFs model.

communication in Industry 5.0, positioning RF–RIS as a promising approach for future 6G-driven smart infrastructure.

III. System Model

The Industry 5.0 smart factory demands wireless infrastructure that is both adaptive and predictive. Conventional indoor connectivity methods based on static planning or ray tracing struggle in dynamic factory environments, where robotic arms and human workers continuously alter the radio landscape, causing frequent blockages and unstable links at mmWave and THz frequencies. RIS provide a promising solution by creating programmable reflections that reshape wireless propagation, but their effectiveness depends on knowing where and how to configure them in real time.

RFs naturally complement RIS by constructing high-fidelity DTs of the indoor space from camera images, capturing both geometry and time-evolving object motion, as shown in Fig. 1. Integrating RFs with RIS enables proactive optimization: the system can anticipate blockages, such as a robotic arm moving into a LoS path, and adjust RIS placement or reflection angles to maintain reliable communication. We operationalize this approach by formulating RIS configuration as a DRL task, where a DRL agent learns optimal placement and phase adjustment strategies from the RF-generated DT. Realistic indoor propagation, including path-loss and small-scale fading models, is incorporated to ensure practical deployability.

IV. Experimental Results

A. Quantitative Evaluation of RF Models

A quantitative evaluation of the visual and structural fidelity of the static DTs was performed by comparing the Splatfacto and Instant-NGP models with the reference factory environment from the Isaac Sim simulator. The evaluation used a set of 248 image pairs, each consisting of a training image and a corresponding image rendered from the trained model at the same camera pose. The final score for each metric represents the mean value across all comparisons.

As reported in Table I, the Splatfacto model significantly outperforms Instant-NGP, achieving higher PSNR and SSIM scores of 26.64 and 0.91 compared to 21.87 and 0.78 for Instant-NGP. The LPIPS metric further confirms this result, with Splatfacto achieving a lower score of 0.15 versus 0.31, indicating superior perceptual similarity. These results demonstrate that Splatfacto provides a more accurate visual and structural representation, making it more suitable for constructing the factory DT.

Table I also presents the quantitative evaluation of the K-Planes dynamic model across RGB, opacity, and disparity images. Trained on 243 robotic arm images, the model's rendered images were compared with the original ground-truth (GT) images from Isaac Sim. The results show that K-Planes effectively reconstructs RGB images with high fidelity, maintains strong structural consistency in opacity images, and produces depth images with a high level of similarity to the original disparity maps.

TABLE I: Average mean values of the static Splatfacto and Instant-NGP models for the entire factory, and of the dynamic K-Planes model for the robotic arm.

Metrics	PSNR ↑	SSIM ↑	LPIPS ↓
Splatfacto	26.64	0.91	0.15
Instant-NGP	21.87	0.78	0.31
K-Planes	29.37	0.97	0.06
K-Planes (*Opacity*)	25.26	0.98	0.04
K-Planes (*Disparity*)	24.56	0.97	0.04

B. Results Obtained from Integrated RF-RIS Model

To evaluate the performance of the proposed PPO-based RIS optimization framework, a comparative analysis is conducted against several benchmark schemes, following the evaluation setup described in Section IV-A. The considered baseline methods are summarized as follows:

- *Random Phase Configuration:* RIS phase shifts are assigned randomly without optimization, serving as a lower-bound performance reference.
- *Static RIS Deployment:* The RIS is installed at a fixed location on the side wall with a static phase configuration.
- *DQN-Based Optimization:* A deep Q-network approach that optimizes RIS phase states using a discrete action space.
- *D3QN-Based Optimization:* An enhanced variant of DQN incorporating double Q-learning and a dueling network architecture to improve convergence stability.
- *DDPG-Based Optimization:* A continuous-action RL method for RIS phase optimization without adaptive placement control.
- *SAC-Based Optimization:* A strong baseline employing the soft actor–critic framework for joint RIS control.
- *Proposed PPO-Based Optimization:* The proposed approach integrates RF-based environmental modeling with

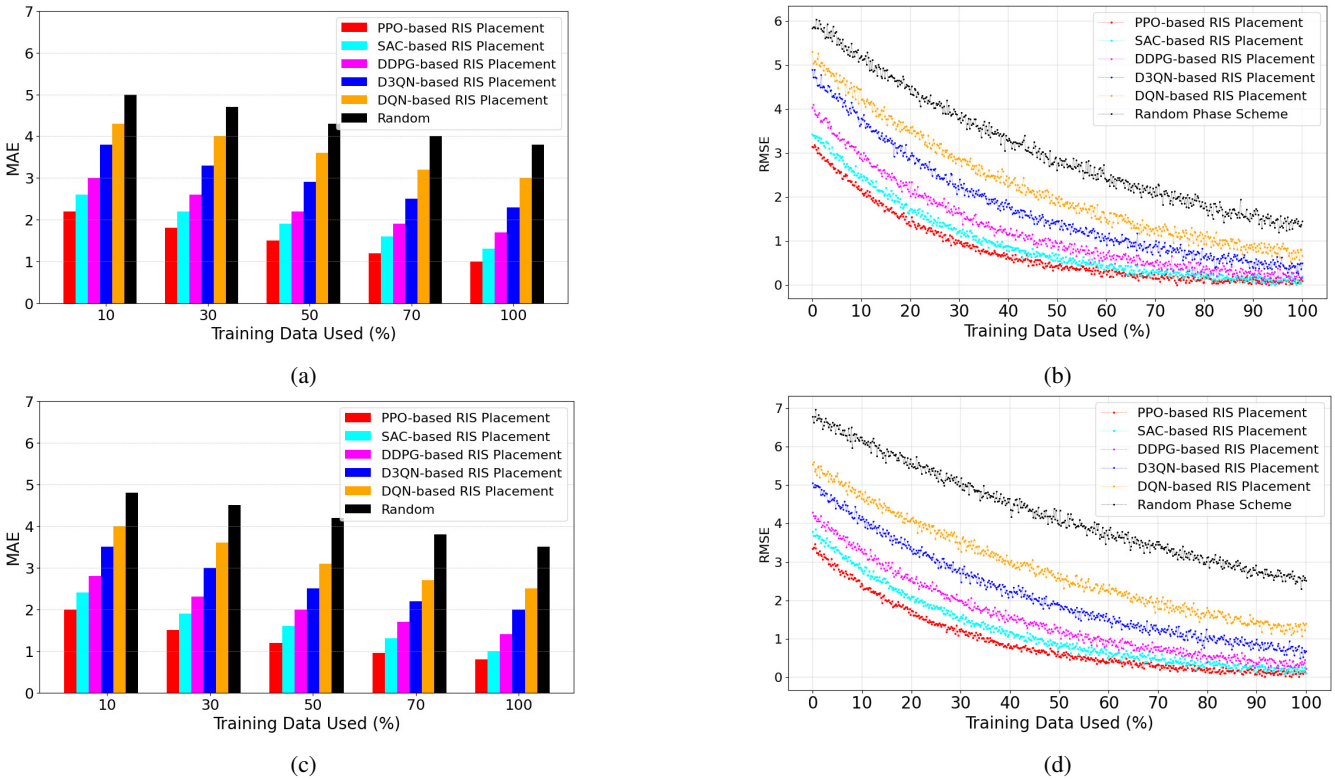


Fig. 2: Performance comparison between our proposed PPO-based optimization and baseline schemes: (a) MAE for Instant-NGP, (b) RMSE for Instant-NGP, (c) MAE for Splatfacto, and (d) RMSE for Splatfacto.

PPO to jointly optimize RIS placement and phase shifts in an adaptive manner.

Fig. 2 presents a performance comparison of the proposed framework and the benchmark schemes using Instant-NGP and Splatfacto models. The subfigures depict the convergence behavior of the mean absolute error (MAE) and root mean square error (RMSE) metrics as the percentage of available training data increases. For both RF models, the proposed PPO-based method consistently achieves lower error levels and faster convergence than all baseline approaches. In particular, Splatfacto demonstrates superior reconstruction accuracy due to its enhanced scene representation capability. Although Instant-NGP exhibits higher absolute error values, it still benefits significantly from adaptive RIS optimization. Overall, lower MAE and RMSE values indicate improved agreement between the estimated and GT wireless channels, leading to more reliable communication performance and robust scene reconstruction.

V. Conclusion

This work presented an integrated framework combining RIS, RFs, and DRL for adaptive wireless connectivity in smart factories. Using RF-based DTs, the system anticipates dynamic layout changes, while the PPO-based optimization jointly adjusts RIS placement and reflection coefficients, significantly reducing MAE and RMSE compared to baselines.

Future extensions could explore cooperative multi-RIS strategies, where signals blocked at one RIS are redirected via another to maintain reliable coverage in smart factories. Such cooperation could also enable adaptive rerouting to robotic arms or other moving devices, ensuring continuous connectivity even in the presence of dynamic obstacles.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic, and the Slovak Academy of Sciences under Grant VEGA 1/0685/23 and by the Slovak Research and Development Agency under Grant APVV SK-CZ-RD-21-0028 and APVV-23-0512.

References

[1] J. Cheng, Y. Yang, X. Zou, and Y. Zuo, "5G in manufacturing: a literature review and future research," *The International J. Advanced Manufacturing Technology*, vol. 131, no. 11, pp. 5637–5659, Apr. 2024.

[2] S. Tripathi, N. V. Sabu, A. K. Gupta, and H. S. Dhillon, *Millimeter-Wave and Terahertz Spectrum for 6G Wireless*. Cham: Springer International Publishing, 2021, pp. 83–121.

[3] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable Intelligent Surfaces: Principles and Opportunities," *IEEE Commun. Surveys & Tuts.*, vol. 23, no. 3, pp. 1546–1577, 2021.

[4] Y. Zhu, B. Mao, and N. Kato, "IRS-Aided High-Accuracy Positioning for Autonomous Driving Toward 6G: A Tutorial," *IEEE Vehicular Technology Magazine*, vol. 19, no. 1, pp. 85–96, 2024.

[5] Sara Fridovich-Keil and Giacomo Meanti, F. R. Warburg, B. Recht, and A. Kanazawa, "K-Planes: Explicit Radiance Fields in Space, Time, and Appearance," in *CVPR*, 2023.

[6] T. Li, H. Lei, H. Guo, M. Yin, Y. Hu, Q. Zhu, and S. Rangan, "Digital Twin-Enhanced Wireless Indoor Navigation: Achieving Efficient Environment Sensing with Zero-Shot Reinforcement Learning," 2024.

[7] S. Alikhani and A. Alkhateeb, "Digital Twin Aided RIS Communication: Robust Beamforming and Interference Management," 2024.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," 2020.

[9] C. Xu, B. Wu, J. Hou, S. Tsai, R. Li, J. Wang, W. Zhan, Z. He, P. Vajda, K. Keutzer, and M. Tomizuka, "NeRF-Det: Learning Geometry-Aware Volumetric Representation for Multi-View 3D Object Detection," 2023.

[10] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, p. 1–15, Jul. 2022.

[11] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," 2023.

Towards Procedural Generation of Degraded Content Using Hierarchical Data Structures

[1]*Peter POPRÍK (2nd year),*
Supervisor: [2]*Branislav MADOŠ*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]peter.poprik@tuke.sk, [2]branislav.mados@tuke.sk

***Abstract*—This paper is concerned with procedural generation of content. Previous work focused on two domain-specific generators for trees and rivers, as well as a bibliometric analysis and a literature review. The research identified the usage of hierarchical data structures in game engines and procedural generation, and a relative lack of research on procedural generation of degraded content. Based on these findings, the paper outlines future plans for a more thorough systematic review and for investigating hierarchical representations as a basis for controllable, structurally plausible generation of degraded content.**

***Keywords*—3D modeling, game design, hierarchical data structures, procedural content generation**

I. Introduction

Procedural content generation (PCG) is a research area that focuses on the algorithmic creation of content. There are many different types of content that can be generated procedurally. Visual content, such as textures or 3D models, is one of the most common types of content generated using PCG. However, PCG can also be used to generate audio content, such as music and sound effects, as well as narrative content, like quests and dialogue, and structural content, such as game levels, maps, rules or mechanics.

Generating different types of content often requires widely different approaches. As a result, the field of PCG is very broad, and existing definitions tend to be formulated in very general terms. While this broadness makes them widely applicable, it also limits their practical usefulness when trying to precisely define the subject of PCG. *Togelius et al.* [1] are concerned with the question of what should be considered PCG, and which properties are essential rather than optional for this definition. They formulate the most generally accepted definition of PCG as "the algorithmic creation of game content with limited or indirect user input". Content, in this case, is defined as all game elements which affect player experience.

PCG can be organized into multiple subfields based on the type of content being generated. For example, procedural modeling focuses on the creation of geometric structures such as buildings, vegetation and terrains, using grammars, noise-based methods or other techniques. Procedural texturing refers to the synthesis and application of textures to 3D models, often using noise functions or other mathematical algorithms. Other subfields concentrate on the construction of game spaces and levels, or the generation of narratives, quests, and puzzles.

Although non-digital content types, like rules for board games can be procedurally generated, and such approaches can be considered some of the earliest forms of PCG, current research in the field is mostly focused on the generation of digital content. Procedural generation is most commonly used in the context of video games, where it refers to the generation of digital *game content*. Because of this, the term is often used interchangeably with procedural content generation for games (PCG-G), which was first defined by *Hendrikx et al.* [2].

II. Motivation

There are several reasons to use procedural techniques for content generation. In modern video game development and other digital media, there is a constant demand for more content. Rising expectations for the quality of content combined with the rising production costs of content have led to a search for more efficient methods of content creation. Instead of manually authoring every asset, designers can specify rules, and let generators produce many variations of similar content with a consistent style and functionality. This makes PCG a useful tool for reducing development time and costs. High budget projects, such as AAA video games, often require large teams of artists and designers producing content and PCG can reduce the amount of manual work needed to create content. On the other hand, smaller studios can use PCG to compete with larger studios despite their limited resources [3].

Another motivation for using PCG, particularly in the context of game design, is its connection to *replayability* [4]. Replayability signifies the potential of a game to be played multiple times or the motivation for a player to keep playing a game for a longer time. This can only be achieved by providing the player with unique and interesting content.

In the context of degraded content, each asset can exist in multiple states of degradation. This additional dimension of possible variations greatly increases the amount of content that would need to be created manually, which makes it a good candidate for procedural generation. Degraded content is discussed in more detail in section IV.

III. Background and previous work

This section summarizes the work carried out in the previous academic year. The research was focused on the development of procedural content generators of trees and rivers, as well as a bibliometric analysis and literature review of the field of PCG.

A. Publications related to procedural content generation

In the previous academic year, research related to the doctoral thesis also resulted in two publications focused on domain-specific procedural generators. The first work [5] proposes an algorithmic generator of three-dimensional polygonal surface models of trees, using a recursive branching scheme with symmetrical and asymmetrical branching modes and a parametrized description of trunk and branch geometry. The generator was implemented as a web application in JavaScript using the Three.js library for visualization and export of models. Experimental evaluation demonstrated that a wide range of diverse tree models can be generated with generation times suitable for interactive use.

The second publication [6] presents a procedural content generator for three-dimensional polygonal models of rivers based on a concept of control tiles with control and connection points that constrain and guide the shape of the river path. The proposed generator separates the process into sketching, modification and rendering phases, allowing efficient computation of river courses while giving control over flow direction, branching and width. Performance tests confirm that river models can be generated in real-time, making it suitable for integration into real-time or interactive environment authoring tools.

B. Bibliometric analysis and literature review

A bibliometric analysis was carried out as part of the work on the doctoral thesis to obtain an overview of research activity in PCG over time. Using the Scopus and Web of Science databases, search queries were constructed that combined general PCG-related terms (such as “procedural generation” and “procedural modelling”) with terms describing various content types, and the results were restricted to publications in computer science, engineering and mathematics between 1970 and 2025. For the purposes of this analysis, all resulting publications were treated as relevant, without further manual filtering. The results show that the number of PCG-related publications remained low until the early 2000s and started to grow noticeably around 2003, with a pronounced gradual increase after 2010, indicating that PCG has recently become a rapidly developing research area.

In addition to the bibliometric analysis, secondary studies like surveys and taxonomies focused on PCG were identified and reviewed. A systematic search in Scopus targeting review-type publications between 2010 and 2025 yielded more than one thousand candidates, from which thirty-seven relevant secondary studies were selected after manual review. These studies cover general PCG overviews, search-based PCG [7], experience-driven and player-adaptive generation [8], PCG via machine learning [9], and a range of content-specific domains such as architecture, terrains, game spaces, textures, sound, stories and puzzles. The distribution of these surveys across content types shows that areas such as architecture, textures and game spaces have a substantial amount of research, whereas stories, puzzles and sound are represented by only a small number of secondary studies. For several content types that are common in games, such as characters, animations, music and typical in-game items, no dedicated surveys were found, which points to potential gaps and opportunities for future research.

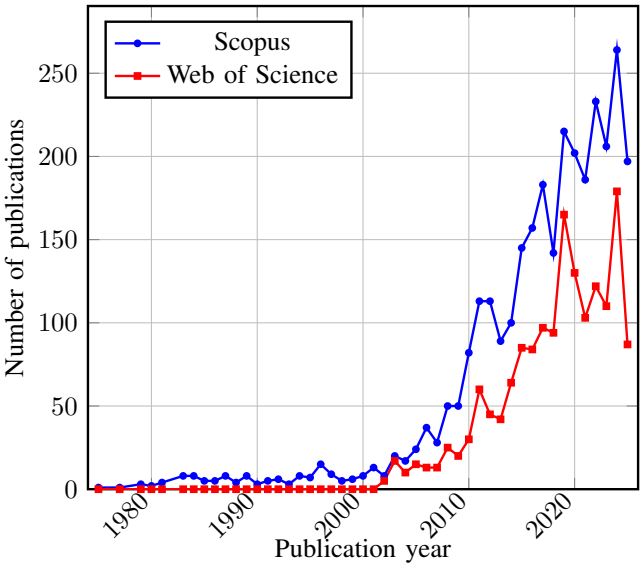


Fig. 1: The number of PCG-related publications per year in the Scopus and Web of Science databases.

There are multiple important findings stated in the reviewed surveys. One of the challenges is creating a universal framework for content generation that can be applied across different content types, as the current methods are often focused on specific types of content [10]. The most commonly mentioned problem is the need to generate high-quality, realistic content. This is a pervasive challenge, since expectations for content quality are always rising [2], [11].

IV. Destruction and Degraded Content Generation

A specific category of content relevant to many games and virtual environments is content that is degraded or damaged in some way. This can include objects that are broken, decayed, or otherwise visually or functionally impaired. The generation of such content is important for a variety of applications. First, a common use case is the generation of destructible environments in games, where players can interact with and damage objects, and the game needs to generate appropriate visual and physical responses to these interactions. Second, degraded content can be used to create more realistic environments, such as abandoned buildings or post-apocalyptic settings, which can only be achieved by using assets with visual degradation. While in both cases the content can be created manually, it would be even more time consuming and costly to create all possible variations than with non-degraded content.

A number of methods for destruction and degradation have been proposed in the computer graphics literature, most of which are based on simulating processes such as fracturing, erosion, corrosion, burning or melting [12], with most of these methods applying destruction to already existing content. In contrast, there are relatively few methods that focus on the procedural generation of degraded content, without relying on existing content.

V. Hierarchical Data Structures in Procedural Content Generation and Game Design

The research conducted and outlined in the previous sections has led to the identification of hierarchical data structures as

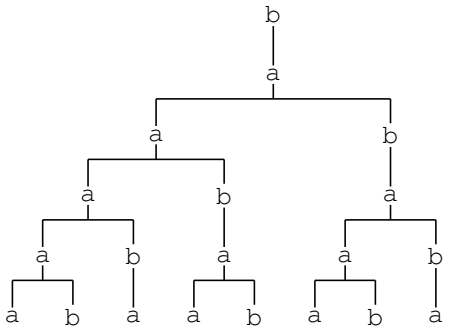


Fig. 2: Derivation tree produced by the parallel rewriting rules $a \to ab$ and $b \to a$. Starting from b, each step replaces all symbols simultaneously [15].

a fundamental method to represent content in the context of procedural generation and game design. Modern game engines such as *Unity* and *Unreal Engine* represent scenes as trees of objects, where parent-child relationships encode logical dependencies between these objects. A scene in a game engine typically consists of a root node representing the entire scene, with subtrees for geometry, dynamic objects, characters, and other elements. Manipulation of a node, such as applying a transformation, will affect all of its children nodes, which allows for efficient control of objects and their components.

Many PCG systems also implicitly or explicitly use hierarchical representations for content generation. Grammar-based methods, like L-systems [13], generate content by recursively applying production rules on symbols, expanding them in a tree-like structure. This process is illustrated in Fig. 2. Similarly, space partitioning methods, such as Binary Space Partitioning (BSP) [14], are used for level generation, where a game space is constructed by recursively subdividing it into smaller regions, creating a tree of partitions.

Generation of degraded content can also benefit from a hierarchical representation. The parent-child relationships in a hierarchical structure can be used to model natural constraints for a degradation process, allowing for the generation of more plausible and structurally logical degradation patterns. Another benefit of a hierarchical organization is more control over the level of degradation, since rules can be applied to specific levels or subtrees of the hierarchy, which is useful for partially degraded content.

VI. Future plans

The next step will be to conduct a more thorough systematic review that combines database searches with backward and forward snowballing on publications, in order to find more methods that may not have been retrieved by the first search. Building on this extended review and on the role of hierarchical data structures outlined in section V, subsequent work will focus on how hierarchical data structures can be used for representation and generation of degraded content. The idea is to investigate how effective can hierarchical dependencies be as a form of constraint, and whether these constraints will allow for the generation of plausible degradation patterns while maintaining control over the process and keeping adequate computational costs.

VII. Conclusion

Procedural content generation is an active research area with applications in game design and other digital media. The use of hierarchical data structures is fundamental for representing content in games and is also commonly used in PCG methods. Generation of degraded content is an important aspect of game design, but there are relatively few methods that focus on procedural generation of such content that are published in academic literature. Future work will focus on investigating how hierarchical data structures can be used for the representation and generation of degraded content, with the aim of generating plausible degradation patterns while maintaining control and managing computational costs.

Acknowledgment

This work was supported by KEGA Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under Grant No. 015TUKE-4/2024 Modern Methods and Education Forms in the Cybersecurity Education.

References

[1] J. Togelius, E. Kastbjerg, D. Schedl, and G. N. Yannakakis, "What is procedural content generation? Mario on the borderline," in *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games*, ser. PCGames '11. New York, NY, USA: Association for Computing Machinery, Jun. 2011, pp. 1–6.

[2] M. Hendrikx, S. Meijer, J. Van Der Velden, and A. Iosup, "Procedural content generation for games: A survey," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, no. 1, pp. 1:1–1:22, Feb. 2013.

[3] J. Togelius, N. Shaker, and M. J. Nelson, "Introduction," in *Procedural Content Generation in Games*, N. Shaker, J. Togelius, and M. J. Nelson, Eds. Cham: Springer International Publishing, 2016, pp. 1–15.

[4] G. Smith, "Understanding procedural content generation: A design-centric analysis of the role of PCG in games," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 917–926.

[5] P. Poprík, B. Madoš, N. Adám, H. Khorshidiyeh, A. Sklenáriková, and L. Tomaščík, "Procedural Content Generator of a Polygonal Surface Models of Trees," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Jan. 2025, pp. 000 203–000 208.

[6] A. Sklenáriková, B. Madoš, N. Ádám, H. Khorshidiyeh, P. Poprík, and L. Tomaščík, "Procedural Content Generator for Generating a 3D Graphical Representation of the River Model," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Jan. 2025, pp. 000 231–000 236.

[7] J. Togelius, G. N. Yannakakis, K. O. Stanley, and C. Browne, "Search-Based Procedural Content Generation: A Taxonomy and Survey," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 3, no. 3, pp. 172–186, Sep. 2011.

[8] G. N. Yannakakis and J. Togelius, "Experience-Driven Procedural Content Generation," *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 147–161, Jul. 2011.

[9] A. Summerville, S. Snodgrass, M. Guzdial, C. Holmgård, A. K. Hoover, A. Isaksen, A. Nealen, and J. Togelius, "Procedural Content Generation via Machine Learning (PCGML)," *IEEE Transactions on Games*, vol. 10, no. 3, pp. 257–270, Sep. 2018.

[10] R. M. Smelik, T. Tutenel, R. Bidarra, and B. Benes, "A Survey on Procedural Modelling for Virtual Worlds," *Computer Graphics Forum*, vol. 33, no. 6, pp. 31–50, 2014.

[11] J. Freiknecht and W. Effelsberg, "A Survey on the Procedural Generation of Virtual Worlds," *Multimodal Technologies and Interaction*, vol. 1, no. 4, p. 27, Dec. 2017.

[12] R. Smelik, F. van Wermeskerken, R. Krijnen, and F. Kuijper, "Dynamic synthetic environments: A survey," *The Journal of Defense Modeling and Simulation*, vol. 16, no. 3, pp. 255–271, Jul. 2019.

[13] A. Lindenmayer, "Mathematical models for cellular interactions in development I. Filaments with one-sided inputs," *Journal of Theoretical Biology*, vol. 18, no. 3, pp. 280–299, Mar. 1968.

[14] N. Shaker, A. Liapis, J. Togelius, R. Lopes, and R. Bidarra, "Constructive generation methods for dungeons and levels," in *Procedural Content Generation in Games*, N. Shaker, J. Togelius, and M. J. Nelson, Eds. Cham: Springer International Publishing, 2016, pp. 31–55.

[15] P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants*, ser. The Virtual Laboratory. New York, NY: Springer, 1990.

Dysarthric Speech Detection in Parkinson's Disease Patients Using Foundation Models and Domain Generalization Techniques

[1]*Ing. Sebastián HREŠKO (1st year),*
Supervisor: [2]*prof. Ing. Liberios Vokorokos PhD.*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]sebastian.hresko@tuke.sk, [2]liberios.vokorokos@tuke.sk

***Abstract*—Dysarthria, a motor speech disorder resulting from neurological impairment, is a key symptom in Parkinson's Disease (PD) and other disorders like ALS and Alzheimer. Automated dysarthria detection has gained significant traction with advances in machine learning. This work presents a comprehensive analysis of data-driven approaches, emphasizing the role of foundation models and domain adaptation strategies for detecting dysarthric speech, especially in PD. We critically analyze state-of-the-art approaches including CNNs, RNNs and self-supervised transformers (e.g., wav2vec 2.0 and Whisper) highlight challenges such as domain shift and data scarcity, and point toward future directions focused on more generalizable and interpretable models.**

***Keywords*—Dysarthria, Parkinson's Disease, Foundation Models, Speech Analysis, Domain Generalization**

I. Introduction

Dysarthria is a motor speech disorder caused by neuromuscular impairments that affect the strength, coordination, and precision of speech muscles, resulting in slurred, monotone, or strained speech that significantly impacts intelligibility and communicative effectiveness. Automated detection of dysarthric speech using machine learning techniques has become a major research focus due to its potential for objective, scalable, and non-invasive assessment of speech impairments. Traditional approaches to dysarthric speech classification employed handcrafted acoustic features such as Mel-frequency cepstral coefficients (MFCCs) and spectral representations, integrated with classical machine learning classifiers, but these methods often struggle with variability in atypical speech patterns [1].

Publicly available dysarthric speech corpora such as UA-Speech [2] and TORGO [3] provide benchmarks for training and evaluating automatic speech recognition (ASR) and dysarthria detection systems, facilitating comparisons among different algorithmic approaches,. However, the limited size and diversity of these datasets present challenges for building models that generalize well to new speakers or recording conditions.

Recent advances in self-supervised and foundation models, including wav2vec2.0 [4] and HuBERT [5], have significantly improved the quality of speech representations by pretraining on large amounts of unlabeled audio. These models can be fine-tuned for downstream tasks such as dysarthric speech detection and severity classification, often outperforming traditional acoustic features, . Transformer-based architectures like Whisper [6], pretrained on extensive multilingual speech-to-text datasets, further enhance robustness to varied speech inputs and noise conditions.

Despite these advancements, substantial challenges remain. Dysarthric speech exhibits high inter-speaker variability, and models frequently encounter performance degradation due to domain shift arising from differences in recording environments, speaker demographics, or language variations. Thus, effective strategies such as domain adaptation, augmentation, and generalization are essential for developing robust systems.

This paper critically analyzes the current state of automated dysarthric speech detection in the context of machine learning, with a focus on foundation models and techniques to improve robustness under domain shift.

II. Literature Review and Analysis

A. Dysarthria and Parkinson's Disease

In Parkinson's Disease (PD), dysarthria is a common speech disorder that affects how people produce sounds and connect words. The form of dysarthria most frequently linked with PD is called hypokinetic dysarthria. This condition appears in most individuals with PD and is characterized by reduced vocal loudness, monotone pitch, imprecise consonant and vowel articulation, and changes in speech rhythm and fluency [7]. These changes can make speech quieter, less clear, and harder to understand.

Hypokinetic dysarthria arises because the disease affects the neural pathways involved in coordinating muscle movements needed for speech. As the disease progresses, these impairments tend to become more pronounced, and objective analyses of speech acoustics have been proposed as a way to detect and monitor changes in speech over time [8]. Studies show that acoustic features such as reduced intensity, monotone voice quality, and altered prosody can be quantified and may help differentiate PD speech from typical speech patterns, making them useful in speech analysis systems.

Since speech abnormalities often occur early in the disease and may precede more obvious motor symptoms, analysing speech acoustics provides a non-invasive method for early detection and long-term monitoring of PD. Understanding the distinctive characteristics of hypokinetic dysarthria and their

measurable acoustic signatures is important when designing machine learning models for automatic detection and tracking of speech changes in PD.

B. Acoustic and Linguistic Biomarkers for PD

PD-related dysarthria is reflected in both segmental attributes, like articulation precision, and suprasegmental properties, such as prosody, timing, and speech rhythm. These changes lead to measurable differences in pitch, intensity, duration, and spectral profiles, which can be extracted using time–frequency representations such as spectrograms and analyzed with signal processing techniques [9].

Common low-level descriptors include Mel-frequency cepstral coefficients (MFCCs), fundamental frequency variations, jitter, shimmer, and harmonics-to-noise ratio (HNR), which capture spectral envelope and perturbations associated with impaired phonation and articulation in PD speech. The rapid repetition of syllables in diadochokinetic tasks (e.g., /pa-ta-ka/) and sustained vowel phonation are widely used in both clinical assessment and computational analysis to isolate specific acoustic patterns and improve disease discrimination performance in machine learning frameworks [10].

More complex combinations of acoustic and linguistic features, such as spectral dynamics over time and measures of speech fluency, have also been explored [11]. These approaches, when integrated with machine learning models, often show strong performance in distinguishing PD speech from control speech and hold promise for objective and scalable early detection and monitoring of the disease.

C. Machine Learning Approaches for Dysarthric Speech

A wide range of machine learning methods have been applied to the task of detecting and classifying dysarthric speech. Early work combined traditional statistical models such as support vector machines (SVMs) and Gaussian mixture models (GMMs) with handcrafted acoustic features, demonstrating the feasibility of automated analysis but often struggling with speaker variability and limited generalization [12].

With the rise of deep learning, architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have become widely adopted for dysarthric speech analysis. CNNs have been shown to effectively learn spectral and temporal patterns from Mel-spectrogram and STFT representations, achieving high performance on benchmark datasets like TORGO and UA-Speech [13]. Hybrid models that integrate CNNs with recurrent units such as long short-term memory (LSTM) layers further exploit both spatial and temporal cues, improving recognition of dysarthric speech [14].

More recently, self-supervised and pretrained models have been explored for dysarthria detection. These models are trained on large amounts of unlabeled speech and can be adapted to pathological speech with transfer learning, often providing richer representations than handcrafted features [15]. For example, SSL models such as wav2vec2.0 and HuBERT have shown promise in dysarthric speech recognition when fine-tuned on impaired speech corpora, although challenges remain due to high phoneme variability in dysarthric signals.

D. Foundation Models in Speech Analysis

Foundation models in speech are large self-supervised or weakly supervised neural networks that learn rich acoustic, phonetic, and prosodic representations from massive amounts of unlabeled or weakly labeled audio. These models serve as powerful feature extractors or end-to-end systems for downstream tasks, requiring minimal task-specific training data.

Self-supervised pretrained models such as wav2vec2.0 have been evaluated as feature extractors for dysarthric speech classification. On the widely used UA-Speech database, embeddings extracted from a pretrained wav2vec2.0 model significantly improved detection and severity classification outcomes compared to traditional spectral and cepstral features. Specifically, features from intermediate layers of wav2vec2.0 produced an absolute detection accuracy improvement of approximately 1.23% over the best performing spectrogram baseline and absolute severity classification improvements of 8.88% and 10.62% compared to MFCC features when using deeper layers of the model [16].

HuBERT (Hidden-Unit BERT) is another self-supervised model that learns latent clusters during pretraining, enabling it to capture higher-level phonetic and acoustic patterns without explicit labels. Comparative studies on dysarthric speech have found that features derived from pretrained HuBERT often outperform wav2vec2.0 features as well as conventional acoustic feature sets such as MFCCs, openSMILE, and eGeMAPS in both automatic detection and severity classification tasks. In experiments on UA-Speech and TORGO databases, HuBERT features achieved consistent performance gains, indicating their stronger representation of dysarthric speech characteristics [15].

End-to-end Whisper models — large encoder-decoder transformers trained on extensive multilingual and multitask speech-to-text corpora — have been applied to dysarthric automatic speech recognition (ASR). In a dysarthric ASR evaluation, Whisper encoder features used with recurrent classifiers such as bidirectional GRUs achieved an identification accuracy of 71.55% on dysarthric speech and demonstrated greater robustness to additive noise compared to wav2vec2.0 based systems [17]. Furthermore, probing analyses of Whisper embeddings found that mid-level encoder layers (e.g., encoder layers 13–15) contain the most informative representations for dysarthria detection and assessment, suggesting that even pretrained layers without extensive fine-tuning can provide effective pathological speech features [18].

Together, these results highlight that foundation models not only offer strong baselines for dysarthria detection and classification but also significantly outperform traditional handcrafted speech features. Their ability to generalize from limited labeled data is especially valuable for pathological speech applications, where annotated corpora are often small and expensive to collect. By leveraging pretrained latent representations, researchers can achieve higher detection accuracy, improved severity classification, and more robust ASR performance in dysarthric speech analysis.

E. Domain Shift in PD Speech Detection

One of the most pressing challenges in deploying machine learning systems for Parkinson's Disease (PD) speech analysis is domain shift. Domain shift occurs when the statistical properties of training and testing data differ, leading to performance degradation in real-world applications. In PD speech, such shifts can arise from differences in recording devices, microphone types, room acoustics, background noise, speaker

demographics, and language or dialect. Models trained on benchmark corpora such as TORGO or UA-Speech often achieve strong internal validation results but may fail to generalize to independent clinical recordings collected under different conditions [19].

A comprehensive cross-dataset evaluation by Hireš et al. further demonstrated that machine learning models trained for PD detection on one speech corpus suffer significant accuracy degradation when evaluated in a leave-one-dataset-out (LODO) setting across multiple independent databases [20]. Their findings confirm that both deep learning and traditional classifiers are highly sensitive to inter-dataset variability, underscoring the severity of domain mismatch in pathological speech analysis.

In addition to technical variability, speaker-specific factors such as age, gender, disease severity, and linguistic background further increase inter- and intra-speaker variability. Temporal fluctuations in speech — influenced by medication state, fatigue, or emotional condition — add another layer of complexity. These factors challenge models that assume stable feature distributions and emphasize the need for domain-aware modeling strategies and robust cross-corpus validation protocols.

F. Domain Adaptation and Generalization Techniques

Machine learning models for Parkinson's Disease (PD) and dysarthric speech tasks often face performance degradation when deployed in real-world settings due to domain shift — a mismatch between the training (source) and testing (target) data distributions. To address this, two complementary strategies have emerged: *domain adaptation* and *domain generalization*.

Domain adaptation assumes access to target domain data (often unlabeled) and adjusts a model trained on a source domain to perform well on that target domain. Domain generalization, in contrast, trains models using one or more source domains so they perform robustly on unseen target domains, without needing target data during training.

Adversarial Domain Adaptation. One widely used domain adaptation technique is adversarial training, where a domain discriminator is trained alongside the task classifier to encourage the feature extractor to learn domain-invariant representations. Wang et al. applied this approach for dysarthric speech detection using UA-Speech and TORGO datasets. Their framework, combining classification loss with domain adversarial training and mutual information minimization, achieved an absolute increase of 22.2% in utterance-level recall and 20.0% in speaker-level accuracy when adapting across datasets [21].

Domain-Invariant Feature Learning. Independent of adversarial methods, domain-invariant representation learning methods aim to learn features that are semantically meaningful but not domain-specific. This includes statistical alignment techniques such as maximum mean discrepancy (MMD) and correlation alignment (CORAL), which explicitly reduce differences in feature distributions. In speech pathology, Hu et al. proposed using acoustic-to-articulatory inversion to generate articulatory features that remain more consistent across domains. This led to improvements in disordered speech recognition when transferring between datasets, highlighting the benefits of invariant physiological representations [22].

Prompt-Based Generalization. Prompt learning is a newer technique from the vision and language domains, where models are conditioned on domain-aware prompts while keeping most of the model weights fixed. DoPrompt, introduced for vision transformers, encodes domain-specific information into learnable prompts, which are adapted to unseen domains at test time [23]. While not yet applied to speech, this approach could inform future dysarthria detection frameworks, especially for adapting foundation models like Whisper or HuBERT across clinics or languages.

Meta-Learning and Cross-Domain Validation. Meta-learning approaches simulate domain shift during training by treating each source domain as a meta-task. This forces the model to generalize across synthetic tasks, improving robustness to actual domain shifts. Chen et al. proposed a meta-learning framework with adversarial domain generalization and cross-domain validation that improves out-of-domain performance without needing target data [24]. These strategies are especially useful when labeled data is limited but multiple source domains are available.

Data Augmentation and Multi-Source Training. Simulating domain variation during training is a practical and effective way to improve robustness. Augmentation methods like pitch shifting, time-stretching, additive noise, and reverberation help expose models to variability that they might encounter in deployment. Multi-dataset training — where models are trained on diverse speech corpora — has also been shown to improve generalization by increasing exposure to different speakers, dialects, and recording setups. [25]

Both domain adaptation and domain generalization are essential to building clinically reliable speech models. While adaptation strategies like adversarial learning allow models to adjust to known target domains, generalization techniques such as meta-learning and invariant feature learning prepare models for deployment in new, unseen environments. As dysarthria detection moves closer to real-world applications, combining these approaches will be key to achieving robust and scalable performance.

III. Problems and Future Directions

Even with significant advancements in the automated detection of dysarthria, there are still various challenges that impede its practical use and wider acceptance in clinical settings. These challenges include limitations in data, fluctuations in speech patterns, and technical restrictions that affect how well models can generalize and be interpreted.

Current Limitations

Data Scarcity and Imbalance: Most publicly available PD and dysarthric speech databases contain insufficient sizes and have an unbalanced representation with a very low number of speakers and a disproportionate amount of severity within these groups. The unavailability of data, along with its unbalanced nature, may result in overfitting, poor generalization and unfairness for model training, especially since the vast majority of datasets are comprised of mild to moderately impaired speech and do not expose the model to all forms of speech impairment that are found in clinical populations.

Furthermore, many databases are linguistically and demographically narrow, limiting the database to one language and/or region and including a limited amount of variability by age, sex, accent and socioeconomic status. Therefore, the broad applicability and fairness of trained systems is impacted

due to limited representativeness in both clinical and real world environments.

Sensitivity to Domain Shift: As discussed in previous sections, machine learning models remain vulnerable to variations in recording devices, environments, and speaker characteristics. Even advanced foundation models may exhibit performance degradation when applied outside the distribution of the training data. This sensitivity limits the reliability of systems in real-world and cross-clinic scenarios.

Limited Interpretability: Deep learning architectures, particularly large neural networks and foundation models, are often difficult to interpret. Although attention maps and latent space visualizations provide some insight, they do not always translate into clinically meaningful explanations. Clinicians require transparent outputs that clearly relate acoustic markers to known speech impairments.

Integration with Clinical Workflows: Most existing systems are developed and evaluated in controlled research settings. Few studies address practical considerations such as real-time processing, user-friendly interfaces, interoperability with electronic health records, or regulatory requirements. Bridging the gap between research prototypes and deployable clinical tools remains a significant challenge.

Research Opportunities

Self-Supervised and Continual Learning: The continued development of self-supervised learning methods offers an opportunity to leverage large amounts of unlabeled speech data. In addition, continual learning frameworks could enable systems to adapt incrementally to new patients and environments while retaining previously learned knowledge. Federated learning, as a privacy-preserving training approach, also fits within this category—allowing models to be updated collaboratively across distributed devices without sharing raw data.

Personalized Modeling: Future systems may benefit from personalization strategies tailored to individual speakers. Techniques such as few-shot learning and speaker-adaptive fine-tuning could allow models to specialize over time with minimal labeled data. These strategies could improve performance for patients with atypical speech profiles or speech changes over time.

Multimodal Fusion: Combining acoustic features with complementary modalities such as facial movement analysis, articulatory measurements, wearable sensor data, or electromyography signals could provide a more comprehensive understanding of motor impairment. Multimodal systems may enhance robustness and capture aspects of PD that are not fully expressed in audio alone.

Robust Evaluation Protocols: There is a need for standardized evaluation frameworks that emphasize cross-corpus validation, out-of-distribution testing, and clinically meaningful outcome measures. Moving beyond single-dataset accuracy toward reproducible and comparable benchmarks will improve reliability and transparency across studies.

Clinical Validation and User-Centered Design: Clinical adoption and the overall usefulness of speech-based medical AI systems depend on increased clinician and patient involvement in their development. Validation of developed tools through user studies, longitudinal trials, and real-world usage ensures that the developed tools meet the actual clinical needs of clinicians and patients, as opposed to being merely technically successful in laboratory settings. This is consistent with wider calls for frameworks in speech-language pathology and clinical AI for assessing the validity, reliability, and clinical relevance of speech measures — as opposed to merely technical measures.

Reproducible research practices, open benchmarks, and shared data sets are also equally important. Shared, open access to corpora and standardization of task evaluations allow for fair comparisons among models; provide an opportunity for independent verification of results; and remove barriers to entry for new researchers, thereby speeding up the pace of research in this area.

IV. Conclusion

This review examined the current state of research on automated detection of dysarthric speech in Parkinson's Disease, with a focus on machine learning methods and foundation models. From handcrafted features to transformer-based embeddings, recent advances have significantly improved model accuracy and robustness. Despite these gains, challenges such as domain shift, limited data, and lack of clinical integration persist. Foundation models offer a promising path forward, particularly when paired with domain adaptation techniques and personalized learning strategies. Future efforts should prioritize generalizability, interpretability, and collaboration with clinical stakeholders to enable real-world deployment of these tools in supporting early diagnosis and disease monitoring in PD.

Acknowledgment

This work was supported by by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00017.

References

[1] Z. Qian, K. Xiao, and C. Yu, "A survey of technologies for automatic dysarthric speech recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, 2023. [Online]. Available: https://doi.org/10.1186/s13636-023-00318-2

[2] S. S. T. Group, "Ua-speech: Benchmark corpus for dysarthric speech recognition," https://speechtechnology.web.illinois.edu/uaspeech/, 2008, accessed: 2026-02.

[3] F. Rudzicz, G. Hirst, and P. Van Lieshout, "Torgo database of dysarthric and control speech," https://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html, 2012, accessed: 2026-02.

[4] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *arXiv preprint arXiv:2006.11477*, 2020. [Online]. Available: https://arxiv.org/abs/2006.11477

[5] W. Hsu, B. Bolte, Y. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, "Hubert: Self-supervised speech representation learning by masked prediction of hidden units," *arXiv preprint arXiv:2106.07447*, 2021. [Online]. Available: https://arxiv.org/abs/2106.07447

[6] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," *arXiv preprint arXiv:2212.04356*, 2022. [Online]. Available: https://arxiv.org/abs/2212.04356

[7] S. Zhai, Q. Cui, D. V. Simmons, and D. J. Surmeier, "Distributed dopaminergic signaling in the basal ganglia and its relationship to motor disability in parkinson's disease," 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095943882300123X

[8] L. Brabenec, J. Mekyska, Z. Galáž, and I. Rektorová, "Speech disorders in parkinson's disease: early diagnostics and effects of medication and brain stimulation," *Journal of Neural Transmission*, vol. 124, no. 3, p. 303–334, 2017. [Online]. Available: https://doi.org/10.1007/s00702-017-1676-0

[9] F. Cao, A. P. Vogel, P. Gharahkhani, and M. E. Renteria, "Speech and language biomarkers for parkinson's disease prediction, early diagnosis and progression," *npj Parkinson's Disease*, vol. 11, p. Article 13, 2025. [Online]. Available: https://doi.org/10.1038/s41531-025-00913-4

[10] H. Ham, B. Jin, K. S. Cha, K. A. Woo, J. H. Shin, and H. Kim, "Syllable-based speech characteristics as potential biomarker for differential diagnosis of parkinson's disease, multiple system atrophy, and cerebellar ataxia," *Journal of Neurology*, vol. 272, no. 8, p. 2011–2024, 2025. [Online]. Available: https://doi.org/10.1007/s00415-025-13352-1

[11] M. Dudek, D. Hemmerling, M. Kaczmarska, J. Stepien, M. Daniol, M. Wodzinski, and M. Wojcik-Pedziwiatr, "Analysis of voice, speech, and language biomarkers of parkinson's disease collected in a mixed reality setting," *Sensors*, vol. 25, no. 8, p. 2405, 2025. [Online]. Available: https://doi.org/10.3390/s25082405

[12] Z. Qian and K. Xiao, "A survey of automatic speech recognition for dysarthric speech," *Electronics*, vol. 12, no. 20, p. 4278, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/20/4278

[13] S. Sajiha, K. Radha, S. Gunnam *et al.*, "Automatic dysarthria detection and severity level assessment using cwt-layered cnn model," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, p. 33, 2024. [Online]. Available: https://link.springer.com/article/10.1186/s13636-024-00357-3

[14] S. R. M. Sekhar, G. Kashyap, A. Bhansali, A. A. A., and K. Singh, "Dysarthric-speech detection using transfer learning with convolutional neural networks," *ICT Express*, vol. 8, no. 1, pp. 61–64, 2022. [Online]. Available: https://doi.org/10.1016/j.icte.2021.07.004

[15] F. Javanmardi, S. R. Kadiri, and P. Alku, "Pre-trained models for detection and severity level classification of dysarthria from speech," *Speech Communication*, vol. 164, p. 103047, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639324000190

[16] F. Javanmardi, S. Tirronen, M. Kodali, S. R. Kadiri, and P. Alku, "Wav2vec-based detection and severity level classification of dysarthria from speech," *arXiv preprint arXiv:2309.14107*, 2023. [Online]. Available: https://arxiv.org/abs/2309.14107

[17] H. P. Japan Bhatt and H. A. Patil, "Noise robust whisper features for dysarthric automatic speech recognition," in *Odyssey: The Speaker and Language Recognition Workshop*, 2024. [Online]. Available: https://www.isca-archive.org/odyssey_2024/bhatt24_odyssey.pdf

[18] Z. Yue, D. Kayande, Z. Cvetkovic, and E. Loweimi, "Probing whisper for dysarthric speech in detection and assessment," *arXiv preprint arXiv:2510.04219*, 2025. [Online]. Available: https://arxiv.org/abs/2510.04219

[19] G. Schu, P. Janbakhshi, and I. Kodrasi, "On using the ua-speech and torgo databases to validate automatic dysarthric speech classification approaches," in *ICASSP 2023 – IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023. [Online]. Available: https://inakodrasi.github.io/publications_pdfs/Schu_ICASSP_2023.pdf

[20] M. Hireš, P. Drotár, N. D. Pah, Q. C. Ngo, and D. K. Kumar, "On the inter-dataset generalisation of machine learning approaches to parkinson's disease detection from voice," *International Journal of Medical Informatics*, vol. 179, p. 105237, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505623002551

[21] D. Wang, L. Deng, Y. T. Yeung, X. Chen, X. Liu, and H. Meng, "Unsupervised domain adaptation for dysarthric speech detection via domain adversarial training and mutual information minimization," *arXiv preprint arXiv:2106.10127*, 2021. [Online]. Available: https://arxiv.org/abs/2106.10127

[22] S. Hu, S. Liu, X. Xie, M. Geng, T. Wang *et al.*, "Exploiting cross-domain acoustic-to-articulatory inverted features for disordered speech recognition," *arXiv preprint arXiv:2203.10274*, 2022. [Online]. Available: https://arxiv.org/abs/2203.10274

[23] Z. Zheng, X. Yue, K. Wang, and Y. You, "Prompt vision transformer for domain generalization," in *arXiv preprint arXiv:2208.08914*, 2022. [Online]. Available: https://arxiv.org/abs/2208.08914

[24] K. Chen, D. Zhuang, and J. M. Chang, "Discriminative adversarial domain generalization with meta-learning based cross-domain validation," *arXiv preprint arXiv:2011.00444*, 2020. [Online]. Available: https://arxiv.org/abs/2011.00444

[25] W.-Z. Leung, M. Cross, A. Ragni, and S. Goetze, "Training data augmentation for dysarthric automatic speech recognition by text-to-dysarthric-speech synthesis," in *Proceedings of Interspeech 2024*. Kos Island, Greece: International Speech Communication Association, 2024, pp. 2494–2498. [Online]. Available: https://www.isca-archive.org/abstracts/Interspeech_2024/papers/1645.html

Sparse Wars - Return of Donoho

[1]*Antónia KOVÁČOVÁ (3rd year)*
Supervisor: [2]Ján ŠALIGA

[1,2]Dept. of Computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]antonia.kovacova@tuke.sk, [2]jan.saliga@tuke.sk

***Abstract*— This paper is devoted to compressed sensing (CS) and describes two methods employing CS. The first method focuses on ECG signal compression by exploiting QRS detection and sparse dictionary learning, achieving high compression ratios (CR) while preserving diagnostically relevant information. The second method addresses early fault detection of the motor by leveraging CS for the motor's impedance characteristics estimation. Experimental results obtained on a single-phase motor indicate that the CS-based measurement method is capable of accurately estimating the impedance characteristics of the motor.**

***Keywords* - compressed sensing, ECG, impedance estimation, sparsity**

I. Introduction

Efficient signal acquisition and compression are essential aspects in the design of many contemporary measurement and diagnostic systems. Traditionally, it was assumed that acquisition systems must satisfy the Shannon–Nyquist sampling criterion, according to which a signal has to be sampled at a frequency at least twice as high as the highest frequency component present in the signal. In 2006, Donoho [1] and, independently, Candès, Romberg, and Tao [2] established the theoretical foundations of compressed sensing (CS), a signal acquisition framework that enables reconstruction of sparse signals from sub-Nyquist measurements. Compressed sensing (CS) is an effective acquisition and reconstruction framework for sparse signals [3], enabling substantially lower sampling frequencies and reduced data transfer rates, as compression is inherently performed during the sampling process [4]. The acquired samples are typically transmitted to the receiving side, where the reconstruction stage takes place with most of the computational complexity [5]. There are many fields where Compressive Sensing (CS) can be applied. CS has been widely used in biomedical signal processing, such as electrocardiograms (ECG) [3], [4], [6], plethysmograms (PPG) [7], heart sounds [8], and also in image processing, for example, MRI [9]. Implementing CS in wireless sensor networks helps reduce computational power consumption at sensor nodes and allows the use of data acquisition systems with lower sampling rates [5]. Diagnostic and measurement systems can also benefit from CS, enabling the use of low-cost acquisition systems and reducing the computational load associated with large datasets, thus allowing faster and more efficient diagnostics [10].

This study builds upon the author's PhD research, progressing from CS theory [12] to structured ECG compression [13], and further advancing these results through adaptive dictionary learning and CS-based impedance diagnostics. In this paper, two methods with implemented CS are presented. The first work [11] described in Section III is a method for ECG signal compression where a dictionary learning process based on ECG signal morphology is involved. This article was also recognized in the 'Publication Activity of PhD Students in 2025' announced by the Rector of the Technical University in Košice. Section IV describes a CS-based measurement method for impedance estimation of the induction motor coils [10] for early fault detection based on impedance characteristic changes. The method was also evaluated through practical experiments to examine its ability to detect impedance changes caused by temperature.

II. Compressed Sensing

The signal sparsity and its sparse representation are the main requirements for introducing CS into signal processing. The signal is considered sparse when most of its coefficients are equal, or close to zero, with few significant signal components. The sparsity of a signal k is defined as the cardinality of its support, i.e., the number of nonzero elements in the signal [14]:

$$\|\mathrm{x}\|_0 := \mathrm{card}(supp(x)) = \mathrm{k} \qquad (1)$$

where $\mathrm{x} \in \mathbb{R}^{\mathrm{n}}$ is a sparse signal and $\|\cdot\|_0$ refers to ℓ_0 norm.

CS is a lossy compression method where a sparse signal is sampled at a lower rate than the Shannon-Nyquist rate, defined as the sampling rate twice the highest frequency present in the signal [5]. The under-sampled signal is reconstructed during decompression using approximation techniques [1]. The CS acquisition of a sparse signal x is defined:

$$\mathrm{y} = \Phi\mathrm{x} = \Phi\Psi\alpha = \mathrm{A}\alpha \qquad (2)$$

where $\mathrm{y} \in \mathbb{R}^{\mathrm{M}}$ is the measurement vector acquired by the sensing matrix $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$, $\Psi \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ is the basis and $\alpha \in \mathbb{R}^{\mathrm{N}}$ is an unknown vector of sparse coefficients that satisfies the equation where $\mathrm{M} \ll \mathrm{N}$.

The goal of CS is to find the coefficients of the vector α subjected to the equation (2). The ideal case aims to find an approximation of the original signal $\hat{\alpha}$ with k non-zero coefficients. This can be formulated mathematically as an ℓ_0 minimization:

$$\hat{\alpha} = min\|\alpha\|_0 \quad subject\ to \quad y = A\alpha \qquad (3)$$

The original CS approach considered random sampling [1]; however, deterministic sampling matrices [8], dynamic or event-driven sampling matrices [6] can also be employed. As basis matrices, orthogonal transformations like the discrete cosine transform (DCT) [8], discrete Fourier transform (DFT), discrete wavelet transform (DWT) [16], or discrete multi-wavelet transform (DMWT) [3] are commonly used.

III. ECG COMPRESSION METHOD

The work [11] presents a novel ECG compression method exploiting a QRS detector for sparse dictionary learning. The ECG signal is first divided into frames with one heart-depolarization cycle. Heart depolarization is the process by which the heart muscle cells undergo an electrical change that leads to contraction [17]. One cycle includes atrial depolarization (P-wave), ventricular depolarization (QRS), and ventricular repolarization (T-wave), shown in the ECG signal in *Fig. 1* from the Physikalisch-Technische Bundesanstalt (PTB) database of ECG signals. The segmentation strategy relies on a QRS detection framework developed during the author's PhD research [18], which enables morphology-consistent framing for sparse dictionary learning. Frames of one heart cycle are then divided according to the positions of R-peaks of QRS complexes.

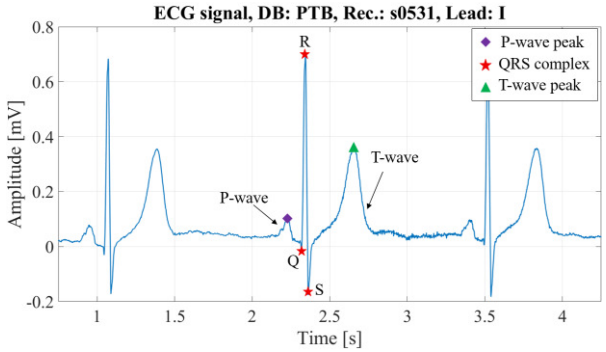


Fig. 1 Segment of an ECG signal from a healthy patient with labelled waves

The goal of sparse dictionary learning is to describe signals with a subset of atoms from the redundant dictionary [19]. The result is a compact signal representation using a parametric dictionary, which can be utilized for signal processing tasks such as denoising and compression. In this work, sparse dictionary learning was employed for dimensionality reduction for signal compression. During this process, signals represented in an orthogonal transform domain are used to reduce the overcomplete dictionary. In this work, the DCT is used to train a sparse dictionary based on a set of training ECG 12-lead signals from the PTB database. The dimensionality reduction is considered as a multiple measurement vectors (MMV) problem, which can be solved via joint minimization. For this purpose, an MMV Orthogonal Matching Pursuit (M-OMP) is utilized for sparse dictionary learning process. As a result, the waveforms of the overcomplete dictionary are reduced to achieve a more compact representation of the ECG signal using its essential DCT components.

The frames of ECG signals are compressed using a deterministic binary block diagonal (DBBD) matrix, which represents the sensing matrix Φ in (2). The compressed and R-peak timestamps are transmitted to the receiving side, where reconstruction of the original signal takes place. The sparse recovery of multiple frames can be jointly performed with MMV sparse recovery [15]. The M-OMP algorithm is employed for reconstruction. Based on the timestamps of R-peaks and their positions in the frames, the recovered frames are merged into one uniform signal. The evaluation was conducted on a set of testing 12-lead records from the PTB database and from the database of the study 'ECG Effects of Dofetilide, Moxifloxacin, Dofetilide-Mexiletine, Dofetilide-Lidocaine and Moxifloxacin-Diltiazem' (ECGDMMLD).

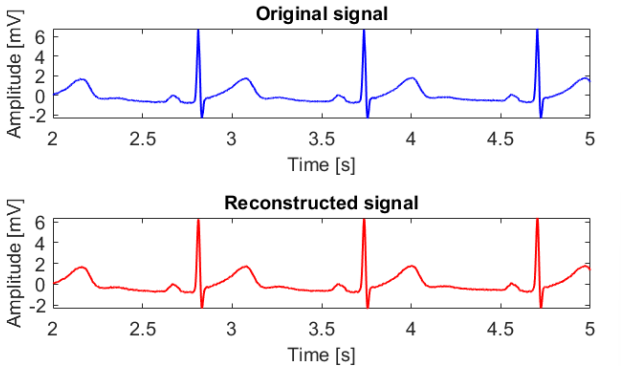


Fig. 2 The comparison of the original ECG signal and reconstructed ECG signal, database ECGDMMLD, set 2001, signal no. 1, lead II.

The analysis of the Percentage Root-mean-square Difference (PRD) and its standard deviation (σ) demonstrated that higher compression ratios (CR ≥ 6) were achieved while preserving crucial medical information (PRD < 9%). The reconstruction errors remained within medically acceptable limits for most leads, for some even at CR = 12. Using a higher threshold during dictionary learning reduces the size of dictionary, which in turn shortens the reconstruction time. The additional metrics, Weighted Diagnostic Distortion (WDD) and the Wavelet Energy–based Diagnostic Distortion (WEDD), were used to evaluate reconstruction quality. The results show WDD indicating very good to good performance up to CR = 12, and WEDD with meeting excellent and very good reconstruction quality.

IV. CS-BASED IMPEDANCE MEASUREMENT

The CS-based impedance measurement method for induction motor coils was presented in work [10]. The method is founded on the assumption that the motor's impulse response is sparse in the time domain. Sparsity enables the reconstruction of the impulse response from a significantly reduced number of samples, allowing impedance estimation with lower sampling rates and reduced data acquisition requirements. The method combines broadband excitation using a Pseudo Random Binary Sequence (PRBS) with CS-based reconstruction and frequency-domain transformation to obtain the impedance spectrum. The motor winding is excited by a PRBS signal, implemented either as a voltage source with a series resistor or, previously, as a current excitation [20], with voltage excitation preferred due to practical current generation limitations. A simplified model of the measurement method is shown in Fig. 3.

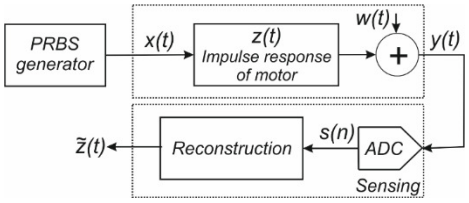


Fig. 3 The CS-based impedance measurement method for an induction motor.

The output voltage $x(t)$ from the PRBS sequence generator is applied to the load under measurement, which in this case is the motor windings. The motor windings behave as a high-frequency system with impulse response $z(t)$, which is convolved with the input voltage and further affected by additive noise $w(t)$. The resulting noisy voltage is then sensed

by the ADC, and a reconstruction algorithm is subsequently applied to obtain the impulse response. Once the impulse response is reconstructed, the Fast Fourier Transform (FFT) is applied to obtain the complex impedance spectrum. The upper frequency limit of the estimated impedance spectrum is defined by the PRBS period. In typical implementations, PRBS periods of 1 µs allow impedance estimation up to several hundred kHz, which is sufficient to capture resonance phenomena associated with parasitic capacitances in industrial induction motors.

The method is evaluated using a high-frequency equivalent circuit model of the induction motor, incorporating parasitic capacitances and resistances. Variations in these parameters, representing insulation degradation or temperature-induced changes, result in observable shifts in resonance frequency and amplitude of the impedance response. Different CS reconstruction algorithms were employed, such as Orthogonal Matching Pursuit (OMP), Iterative Hard Thresholding (IHT), Compressive Sampling Matching Pursuit (CoSaMP), and semidefinite programming-based approaches, such as the SDPT3 solver for ℓ_1 minimization.

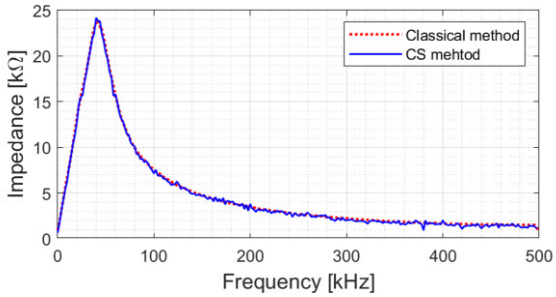


Fig. 4 Comparison of impedance frequency response estimated with SDPT3 algorithm (CR = 50) and standard sampling (Classical method).

Experimental validation of the method confirms that accurate impedance estimation is achievable even at high compression ratios (Fig. 4). The method exhibits robustness against measurement noise, PRBS seed variations, and temporal misalignment between excitation and acquisition. The experiment conducted in a temperature-controlled environment demonstrated a clear linear relationship between temperature and the maximum value of the impedance frequency response.

V. Conclusion

This work summarizes the principal methodological contributions achieved during the author's PhD study, presenting two independently validated CS-based frameworks. For the ECG compression method, the obtained results confirm that the proposed approach achieves high compression ratios while preserving diagnostically relevant information. The PRD-based evaluation and additional distortion metrics indicate that reconstruction quality remains within medically acceptable limits even at high compression ratios. Furthermore, the optimized DCT-based dictionary reduces reconstruction time, indicating improved suitability for real-time applications.

The second method, focused on the estimation of motor impedance characteristics, confirms that accurate impedance estimation can be achieved using CS, even at high compression ratios. The robustness of the proposed method was evaluated in a temperature-varying environment, demonstrating sensitivity to parameter changes. The observed variations in the impedance frequency response indicate that the method is suitable for early fault detection in induction motors. These properties, combined with low hardware complexity, make the CS-based impedance measurement method a practical solution for induction motor diagnostics and condition monitoring.

Acknowledgment

The work is part of a project supported by the Cultural and Educational Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic (No. 025TUKE-4/2026).

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006, doi: 10.1109/TIT.2006.871582.
[2] E. Candes, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," 2005, *arXiv*. doi: 10.48550/ARXIV.MATH/0503066.
[3] O. Kováč, J. Kromka, J. Šaliga, and A. Jusková, "Multiwavelet-based ECG compressed sensing," *Measurement*, vol. 220, p. 113393, Oct. 2023, doi: 10.1016/j.measurement.2023.113393.
[4] J. Šaliga, I. Andráš, P. Dolinský, L. Michaeli, O. Kováč, and J. Kromka, "ECG compressed sensing method" *Measurement*, vol. 183, p. 109803, Oct. 2021, doi: 10.1016/j.measurement.2021.109803.
[5] I. Andráš, P. Dolinský, L. Michaeli, and J. Šaliga, "Sparse Signal Acquisition via Compressed Sensing and Principal Component Analysis," *Meas. Sci. Rev.*, vol. 18, no. 5, pp. 175–182, Oct. 2018, doi: 10.1515/msr-2018-0025.
[6] P. Daponte, L. De Vito, G. Iadarola, and F. Picariello, "ECG Monitoring Based on Dynamic Compressed Sensing of Multi-Lead Signals," *Sensors*, vol. 21, no. 21, p. 7003, Oct. 2021, doi: 10.3390/s21217003.
[7] S. S. Abdulkader and U. A. Qidwai, "A Review on PPG Compression Techniques and Implementations," in *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Langkawi Island, Malaysia: IEEE, Mar. 2021, pp. 511–516. doi: 10.1109/IECBES48179.2021.9398812.
[8] P. Daponte, L. De Vito, G. Iadarola, F. Picariello, and S. Rapuano, "Deterministic Compressed Sensing of heart sound signals," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Lausanne, Switzerland: IEEE, Jun. 2021, pp. 1–6. doi: 10.1109/MeMeA52024.2021.9478766.
[9] M. Sandilya and S. R. Nirmala, "Compressed sensing trends in magnetic resonance imaging," *Eng. Sci. Technol. Int. J.*, vol. 20, no. 4, pp. 1342–1352, Aug. 2017, doi: 10.1016/j.jestch.2017.07.001.
[10] O. Kovac *et al.*, "A CS-based measurement method for impedance estimation of induction motor coil," *Measurement*, vol. 259, p. 119586, Feb. 2026, doi: 10.1016/j.measurement.2025.119586.
[11] A. Kovacova, G. Iadarola, L. D. Vito, O. Kovac, J. Saliga, and J. Sevec, "An ECG compression method exploiting a QRS detector for sparse dictionary learning," *Measurement*, vol. 258, p. 119177, Jan. 2026, doi: 10.1016/j.measurement.2025.119177.
[12] A. Jusková, "Sparse wars: A new compression paradigm," in *24th Scientific Conference of Young Researchers Proceedings from Conference*, Technical University of Kosice, 2024, pp. 87–90.
[13] K. Antónia, "Sparse Wars: Nyquist Strikes Back," in *25th Scientific Conference of Young Researchers Proceedings from Conference*, Technical University of Kosice, 2025, pp. 32–33.
[14] M. Taghouti, "Compressed sensing," in *Computing in Communication Networks*, Elsevier, 2020, pp. 197–215. doi: 10.1016/B978-0-12-820488-7.00023-2.
[15] M. E. Erkoc and N. Karaboga, "A comparative study of multi-objective optimization algorithms for sparse signal reconstruction," *Artif. Intell. Rev.*, vol. 55, no. 4, pp. 3153–3181, Apr. 2022, doi: 10.1007/s10462-021-10073-5.
[16] L. De Vito, E. Picariello, F. Picariello, S. Rapuano, and I. Tudosa, "A Dictionary Optimization Method for Reconstruction of ECG Signals after Compressed Sensing," *Sensors*, vol. 21, no. 16, p. 5282, Aug. 2021, doi: 10.3390/s21165282.
[17] M. J. Mc Loughlin and J. M. Di Diego, "Review on new ECG-derived leads and regional vectorcardiograms," *J. Electrocardiol.*, vol. 81, pp. 13–19, Nov. 2023, doi: 10.1016/j.jelectrocard.2023.07.004.
[18] A. Kováčová, O. Kováč, and J. Šaliga, "Adjustment Algorithm for Multi-Lead QRS Detection," in *2025 15th International Conference on Measurement*, Smolenice, Slovakia: IEEE, Jun. 2025, pp. 23–26. doi: 10.23919/MEASUREMENT66999.2025.11078714.
[19] I. Tosic and P. Frossard, "Dictionary Learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011, doi: 10.1109/MSP.2010.939537.
[20] J. Kromka *et al.*, "Online impedance estimation of induction motor coils using a CS-based measurement method," *Meas. Sens.*, p. 101425, Dec. 2024, doi: 10.1016/j.measen.2024.101425.

Towards Affordable Wearable 3D LiDAR Navigation Assistance: Scanner Design and Calibration

[1]*Richard KRÁL (2nd year)*
Supervisor: [2]Tibor VINCE

[1,2]Dept. of Theoretical and Industrial Engineering, FEI, Technical University of Košice, Slovak Republic

[1]richard.kral@tuke.sk, [2]tibor.vince@tuke.sk

***Abstract*— This paper summarizes research progress toward a wearable 3D LiDAR scanning concept for navigation assistance for people with visual impairment. Existing assistive solutions often remain limited by high cost, computational demands, and insufficient ranging precision of low-cost sensing modalities. To address these barriers, a compact single-point LiDAR is combined with a dual-axis oscillating beam-steering mechanism based on permanent magnets, excitation coils, an H-bridge driver, and torsion springs. Hall-effect probes are used to estimate instantaneous oscillator deflection, and a geometry-referenced calibration procedure is applied to map Hall ADC readings to tilt angle under oblique-range acquisition, including hysteresis-aware up/down handling and robust aggregation to suppress noise and outliers. Proof-of-concept 3D point-cloud visualization is demonstrated on representative stable measurement intervals, capturing indoor structures such as a doorway opening and the adjacent room wall. Future work is directed toward improving calibration robustness during simultaneous dual-axis operation, defining an effective user-feedback channel, integrating the scanning module into a chest-mounted waist-bag form factor, and conducting staged validation including tests with users with visual impairment.**

***Keywords*— wearable 3D LiDAR, assistive navigation, visual impairment, beam steering, Hall-effect calibration, point-cloud reconstruction**

I. Introduction and Motivation

Independent mobility and safe navigation remain essential for people with visual impairment, yet many wearable assistive concepts still face barriers in everyday deployment. In recent research, sensing is increasingly embedded into practical form factors (e.g., garments or cane-based platforms) and combined with multi-sensor architectures to support obstacle avoidance and context-aware assistance [1]–[3], while head-worn "smart-glasses" concepts have also been demonstrated [4]. In parallel, several commercial systems indicate current deployment trends, typically relying on camera-based perception with audio and/or haptic feedback channels (e.g., *Lumen, Sixth Sense, SuperBrain*, and *Niira*) [5]–[8]. However, recurring limitations are still observed, particularly high cost, usability complexity, and sensing constraints in challenging environments.

For distance sensing in assistive navigation, LiDAR-based ranging is attractive due to its direct and repeatable depth measurements and comparatively modest processing requirements relative to stereo vision, while higher spatial precision than low-cost ultrasound is typically achievable. Recent compact single-point LiDAR modules demonstrate that wearable integration can be approached with low mass and power while maintaining obstacle-relevant ranges (e.g., *TFmini Plus* and *TFS20-L)* [9], [10]. At the same time, alternative depth sensors and scanning solutions exist (e.g., ToF arrays such as *VL53L8CX*, 2D LiDAR *LD19*, and 3D LiDAR units such as *VLP-16*), but these are commonly associated with higher cost, size, and power demand when extended to dense 2D/3D perception [11]–[13]. Therefore, an affordable and energy-efficient approach to wearable 3D environment sensing is motivated, in which a compact single-point LiDAR is combined with mechanical beam steering to enable spatial reconstruction while preserving wearability.

II. Background and Progress

A. System concept and hardware architecture

A wearable 3D environment-sensing concept is based on a compact single-point LiDAR module combined with mechanical beam steering. Instead of employing bulky 2D/3D LiDAR units, angular scanning is generated by a rotary oscillator driven electromagnetically. The oscillation principle is illustrated in Fig. 1, where permanent magnets and excitation coils are actuated via an H-bridge driver, while torsion springs provide the restoring torque required for resonant operation. This approach enables a lightweight scanning mechanism and allows the angular motion to be monitored by analog Hall-effect probes that measure magnetic-field intensity and thereby infer instantaneous deflection.

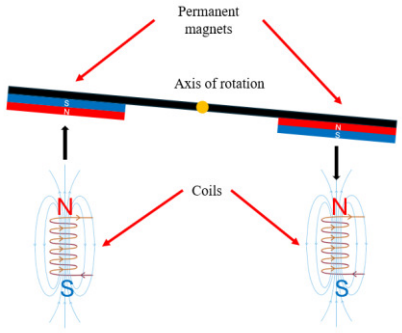


Fig. 1 Oscillation principle of a rotary scanner axis using permanent magnets, excitation coils, and an H-bridge driver.

The scanning mechanism provides angular deflection in two orthogonal axes corresponding to horizontal and vertical directions. The structure consists of two coupled oscillatory subsystems enabling independent motion in each axis; each axis incorporates a pair of permanent magnets with excitation coils, while torsion springs define the mechanical stiffness and natural frequency. In the current implementation, the horizontal and vertical oscillations operate at approximately 14 Hz and 3 Hz, respectively, producing a Lissajous scanning pattern with dense spatial coverage. This frequency combination was determined experimentally to balance power consumption and achievable angular deflection. The oscillation amplitudes reach approximately ±19° in the vertical axis and ±33° in the horizontal axis, resulting in fields of view of about 38° and 66°. Integrating the LiDAR sensor into the scanning unit allows the measurement direction to follow the oscillatory motion, enabling sequential sampling of spatial directions without requiring complex multi-beam or rotating LiDAR systems.

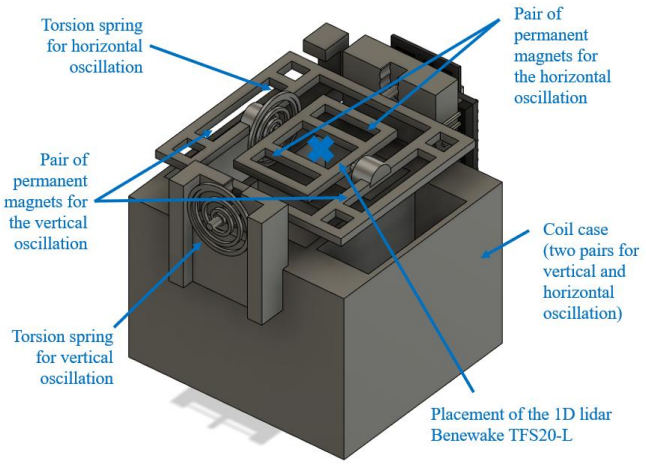


Fig. 2 Dual-axis concept of a wearable 3D LiDAR scanner based on two rotary oscillators and torsion springs.

B. Calibration baseline and current status

Angle estimation is based on converting Hall-sensor ADC readings to instantaneous angular deflection, where accurate mapping is essential for consistent 3D reconstruction. A single-axis calibration pipeline has been established, including (i) construction of a geometric reference deflection from LiDAR measurements to a planar target, (ii) compensation of an approximately constant phase offset between the Hall waveform and the reference signal, and (iii) hysteresis handling by separating increasing and decreasing motion branches. This baseline provides stable angle recovery when axes are operated independently and forms the starting point for further extension to simultaneous scanning.

C. Dual-axis oscillation and measurement implications

Simultaneous excitation of both axes has been achieved in the current prototype, allowing concurrent horizontal and vertical motion to be produced as required for 3D scanning. However, the transition from single-axis operation to concurrent dual-axis oscillation introduces additional effects that are not captured when each axis is calibrated in isolation.

Activation of the vertical oscillator influences the horizontal axis through changes in effective oscillation amplitude and through additional disturbance components observable in the Hall-sensor signal. These effects can be attributed to mechanical coupling, magnetic-field interaction, and resonance shifts under dual excitation. As a result, calibration parameters obtained under single-axis conditions cannot be assumed fully valid during simultaneous two-axis scanning, and increased error is observed in specific phases of the oscillation cycle. Consequently, recalibration and compensation strategies tailored to concurrent two-axis oscillation are required to achieve stable 3D reconstruction over entire measurement intervals.

III. *RECENT ACHIEVEMENTS*

A. Hall-based angle calibration and hysteresis handling

A calibration procedure was established to convert Hall-sensor ADC readings into instantaneous angular deflection of the rotary oscillator. Reference deflection was derived from LiDAR ranging to a planar target using measurement geometry, where the measured slant range $r(t)$ and the known perpendicular distance D were used to compute a time-varying angle $\theta_{\mathrm{ref}}(t)$, while the zero-deflection offset α_0 was estimated robustly from the recorded data by nonlinear least-squares fitting. Because the Hall waveform and $\theta_{\mathrm{ref}}(t)$ were not precisely time-aligned, an approximately constant phase offset was estimated and compensated using sub-sample shifting via interpolation.

To address the direction-dependent behavior of the magnet–Hall sensing chain, the calibration mapping was not treated as a single global function. Instead, separate mappings were constructed for the increasing and decreasing branches of the oscillation (up/down), which reduced hysteresis-related bias and eliminated phase-dependent ambiguity. Robust estimation was supported by aggregating samples across multiple periods and by computing representative values (median) within narrow ADC intervals, thereby attenuating noise and sporadic outliers. A representative calibration outcome is illustrated in Fig. 3, demonstrating stable angle recovery over a selected interval suitable for subsequent spatial reconstruction.

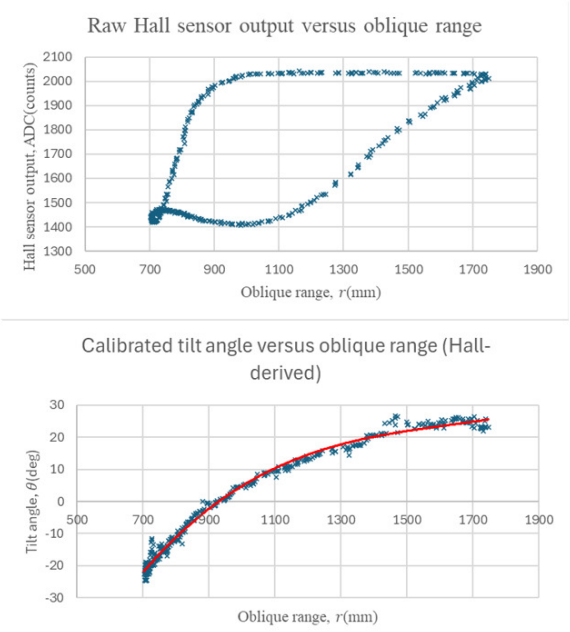


Fig. 3 Raw and calibrated sensor relationships for oblique-range acquisition on the vertical oscillation axis (up: Hall ADC code vs. range; down: angle recovery- Hall-derived tilt angle vs. range).

B. Proof-of-concept 3D reconstruction and visualization

A proof-of-concept 3D reconstruction pipeline was implemented by combining LiDAR range samples with time-resolved beam direction obtained from the oscillator deflection signals. For each measurement instant, a 3D point coordinates was computed from the measured range together with the corresponding steering angles, enabling conversion from polar measurements to Cartesian coordinates and accumulation into a point cloud using custom Python-based application. Basic filtering was applied at PoC level to improve stability of the reconstructed scene, including outlier suppression through robust statistics and temporal/interval aggregation prior to visualization.

During simultaneous dual-axis operation, calibration stability is not uniformly preserved over the entire acquisition; therefore, representative stable intervals are used for proof-of-concept visualization. An example point-cloud reconstruction is shown in Fig. 4, illustrating that the scanning concept can provide structured spatial returns suitable for further refinement, including improved dual-axis calibration consistency, more systematic denoising, and quantitative evaluation in realistic navigation scenarios.

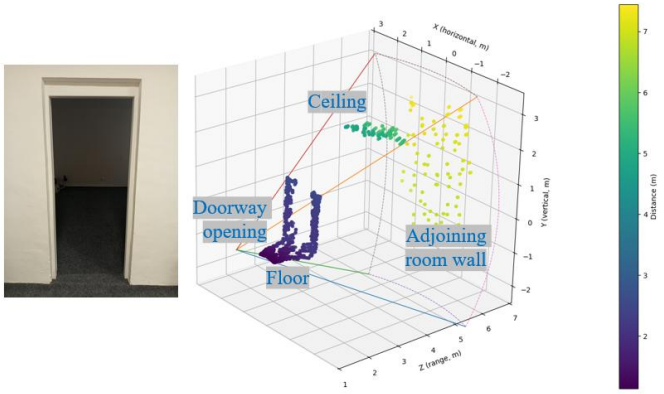


Fig. 4 Proof-of-concept 3D point cloud showing a wall discontinuity corresponding to a doorway opening and the wall of the adjoining room.

IV. Conclusion and Future Plans

An affordable and energy-efficient concept for wearable 3D environment sensing has been advanced, combining a compact single-point LiDAR with a dual-axis oscillating beam-steering mechanism. A Hall-sensor–based angle calibration pipeline has been established and validated for oblique-range acquisition on a single oscillation axis, and proof-of-concept 3D point-cloud visualization has been demonstrated on representative stable data intervals.

The primary technical objective for the forthcoming period is to improve calibration robustness during simultaneous dual-axis operation, where cross-axis interaction and sensor-signal disturbances can reduce mapping consistency. Calibration and compensation strategies tailored to concurrent oscillation are to be developed, including improved phase tracking, branch/segment handling, and enhanced outlier suppression to achieve stable reconstruction over full acquisitions.

For near-term wearable deployment, the current scanning module is planned to be integrated into a chest-mounted waist bag (Fig. 5) that provides sufficient internal volume while maintaining unobtrusive placement and a forward-looking sensing direction aligned with torso orientation. Basic mechanical robustness against everyday shocks and impacts will be considered in the enclosure design. Additionally, an onboard accelerometer will be used to estimate orientation with respect to gravity, ensuring consistent alignment of the sensing frame.

In parallel, an appropriate user-feedback concept is to be designed to translate spatial measurements into actionable mobility cues, including selection of modality (haptic and/or audio), body placement, and an interpretable encoding of direction, distance, and urgency with low cognitive load.

Finally, systematic validation is planned through staged testing that progresses from controlled indoor trials to real-world scenarios. Evaluation is intended to cover navigation safety and efficiency (e.g., collision/near-miss rate and task completion time), together with user workload and acceptance measures, providing evidence for iterative refinement and broader field validation.

Fig. 5 Proposed chest-mounted waist-bag enclosure for the wearable 3D LiDAR prototype

References

[1] S. Bhatlawande, R. Borse, A. Solanke and S. Shilaskar, "A Smart Clothing Approach for Augmenting Mobility of Visually Impaired People," in IEEE Access, vol. 12, pp. 24659-24671, 2024, doi: 10.1109/ACCESS.2024.3364915.

[2] A. I. Apu, A.-A. Nayan, J. Ferdaous, and M. G. Kibria, “IoT-Based Smart Blind Stick,” Lecture Notes on Data Engineering and Communications Technologies. Springer Singapore, pp. 447–460, Dec. 04, 2021. doi: 10.1007/978-981-16-6636-0_34.

[3] K. Minatani, “Proposal for a Versatile Smart White Cane Infrastructure Using a 3D Printer: As a R&D Platform and a Practical Product Design,” Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, pp. 21–26, Jun. 26, 2024. doi: 10.1145/3652037.3652064.

[4] Y.-L. Poy, S. Darmaraju, C.-H. Goh, and B.-H. Kwan, “Standalone Smart Glass System for the Blind and Visually Impaired,” 2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, pp. 239–244, May 24, 2024. doi: 10.1109/iscaie61308.2024.10576410.

[5] Lumen, .Lumen, [quoted 8.1.2026], available on the internet: <https://www.dotlumen.com/>

[6] HopeTech, Sixth sense, [quoted 9.1.2026], available on the internet: <https://www.hopetech.vision/>

[7] 7sense, SuperBrain, [quoted 9.1.2026], available on the internet: < https://7sense.ee/toode/superbrain-i/>

[8] Sensotec and Eyesynth, Niira, [quoted 10.01.2026], available on the internet: <https://www.pro.sensotec.com/post/niira-glasses-launch>

[9] Benewake TFmini Plus (1D LiDAR), [quoted 10.01.2026], available on the internet: <https://en.benewake.com/TFminiPlus/index.html>

[10] Benewake TFS20-L (1D LiDAR, dToF), [quoted 10.01.2026], available on the internet: <https://en.benewake.com/TFS20L/index.html>

[11] ST VL53L8CX (8×8 ToF array), [quoted 11.01.2026], available on the internet: <https://www.st.com/en/imaging-and-photonics-solutions/vl53l8cx.html>

[12] DFROBOT LD19 (2D LiDAR), [quoted 12.01.2026], available on the internet: < https://www.dfrobot.com/product-2610.html?srsltid=AfmBOoqpT4viWRq0Gg-lE3rmNZIq3QbK1kX1-h1u5946pcBKi6A1ufry>

[13] Velodyne VLP-16 (3D LiDAR), [quoted 12.01.2026], available on the internet: < https://www.mapix.com/lidar-scanner-sensors/velodyne/velodyne-puck-vlp16/

Possibilities of using a digital twin for flying shears

[1]*Matej HRIC (2nd year)*
Supervisor: [2]František ĎUROVSKÝ

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Košice, Slovak Republic

[1]matej.hric@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*- Flying shears used in material processing lines require high dynamic performance of the drive system during the cutting phase. Changes in cutting conditions cause variations in drive load, which are not explicitly monitored in standard control systems. Digital twin is considered a virtual representation of the drive system, whose outputs are compared with measured signals such as speed, torque, and current. Differences between nominal and actual drive behavior during cutting can be evaluated. The proposed approach provides a basis for condition monitoring and predictive maintenance of flying shears.**

***Keywords*-Flying shears, Predictive maintenance, Industrial automation, Digital twin**

I. INTRODUCTION

Flying shears are widely used in material processing lines for continuous cutting of moving material. Their operation requires precise synchronization between the material speed and the motion of the cutting mechanism. During the cutting phase, the drive system is exposed to high dynamic loads, which place significant demands on control performance, drive sizing, and mechanical robustness.

During long-term operation, cutting conditions naturally change due to variations in material properties or the gradual degradation of cutting blades. These changes influence the mechanical forces acting on the cutting mechanism and are reflected in drive parameters such as cutting torque, motor current, and speed. While these signals are available in most industrial drive systems, they are typically used only for control and scheduled maintenance purposes and are not further analyzed.

Digital Twin technology represents a promising approach for advanced monitoring of mechatronic systems [1]. A digital twin can be understood as a virtual representation of a physical system that describes its expected dynamic behavior and enables systematic comparison with measured data from the real system [2]-[4]. In the case of flying shears, such an approach allows observation of deviations in drive behavior that occur during the cutting phase and may indicate changes in operating conditions.

This paper focuses on the possibilities of using a digital twin for monitoring the dynamic behavior of a flying shears drive, since there are not that many publications aimed at individual sections of material processing lines [5],[6]. The aim of this paper is not to develop a physical model of blade wear, but to investigate how blade wear influences the dynamic behavior of the drive system during the cutting process. The focus is on identifying measurable indicators derived from changes between the digital twin and real system signals that reflect changes in cutting conditions.

II. FLYING SHEARS AND DRIVE LOAD CHARACTERISTICS

During one operating cycle, the drive accelerates the cutting drums to match the material speed, performs the cutting action, and then decelerates back to the initial state. This results in strong short- duration loads acting on the drive system [7].

The most critical part of the cycle is the cutting phase, during which the cutting resistance causes a significant increase in load torque. This load increase leads to higher torque demand and motor current and may result in a temporary speed drop, depending on the control structure and available kinetic energy of the system. The duration and shape of speed drop depend on cutting conditions, material properties, and the mechanical state of the cutting blades.

From the drive perspective, changes in cutting conditions are observable in electrical and mechanical signals such as cutting torque, current, and angular speed. As cutting torque increases, the dynamic response of the drive during the cutting phase is worse than its nominal behavior. Although these changes contain valuable information about the size of cutting torque or quality of the cutting process, conventional control systems typically consider the operation acceptable if tracking performance remains within predefined limits.

Understanding the relationship between cutting conditions and drive load characteristics is essential for advanced monitoring concepts. The dynamic behavior of the drive during cutting provides information that can be analyzed by a digital twin for further analysis.

III. DIGITAL TWIN CONCEPT FOR FLYING SHEARS DRIVE

The digital twin of flying shears can be considered at the drive system level, with a focus on the dynamic behavior of the electric drive during the operating cycle. In this concept, the digital twin represents a virtual model of the drive and its control structure, describing the nominal dynamic response of the system under defined operating conditions.

Model-based digital twin can be developed using simulation tools commonly applied in industrial drive system design. Environments such as MATLAB/Simulink allow the creation of dynamic models suitable for analyzing load behavior during transient operating states. For future

extensions, specialized tools such as Siemens DriveSim Designer may also be considered, as they provide drive representations compatible with industrial configurations (Fig. *1*),[8].

Siemens DriveSim Designer [9] can be used for modeling and virtual commissioning of electric drive systems. The software enables detailed representation of the drive, including the electric machine, power electronics, mechanical load, and control structure. It provides the expected time evolution of important variables such as torque, current, and speed under both steady-state and transient operating conditions. By comparing simulated responses with measured signals from a real drive system, DriveSim Designer supports validation of drive behavior and model accuracy.

IV. Possibilities of Using Digital Twin for Flying Shears

The operation of flying shears is characterized by highly dynamic mechanical loads, especially during the cutting phase. Changes in cutting conditions or mechanical state directly influence the drive´s behavior. Digital twin enables these effects to be systematically observed and interpreted. The following subsections outline selected possibilities of using a digital twin in flying shears.

A. Monitoring of drive load changes during cutting

A fundamental problem in flying shears operation is the variability of mechanical load during the cutting phase. Increased cutting resistance leads to higher torque demand and current peaks.

As shown in Fig. *2*, the size of the cutting torque is in some cases almost equal to the nominal torque of the drive. For that reason, compensation must be added to the motor to make a cut. By comparing simulated nominal responses with measured drive signals, a digital twin enables observation of these load changes. Variations in torque, current, or short-term speed changes provide information about instantaneous mechanical loads and allow distinction between nominal and critical cutting conditions.

B. Influence of blade wear on dynamic behavior

Blade wear represents a long-term degradation phenomenon that significantly influences the cutting process and the dynamic behavior of flying shears. Unlike short-term variations caused by material or transient operating conditions, blade wear develops gradually over many cutting cycles and leads to systematic changes in mechanical resistance during cutting.

Fig. *3* shows one cutting cycle for speed of 20 m/min, PI speed control with compensation, and 20% wear of blades, which was considered as nominal in this simulation. The speed drop was around 0.6 m/min. As the cutting edges deteriorate, the effective sharpness of the blades decreases, resulting in increased cutting forces and higher mechanical resistance acting on the cutting mechanism. This increase in resistance directly affects the drive system, leading to higher required cutting torque and possible damage to material.

From the perspective of the drive system, blade wear does not show as critical fault but rather as a slow and continuous change in dynamic load characteristics. While the speed controller may still maintain acceptable tracking performance, the effort required to do so increases over time. This results in higher stress on both electrical and mechanical components, potentially reducing efficiency.

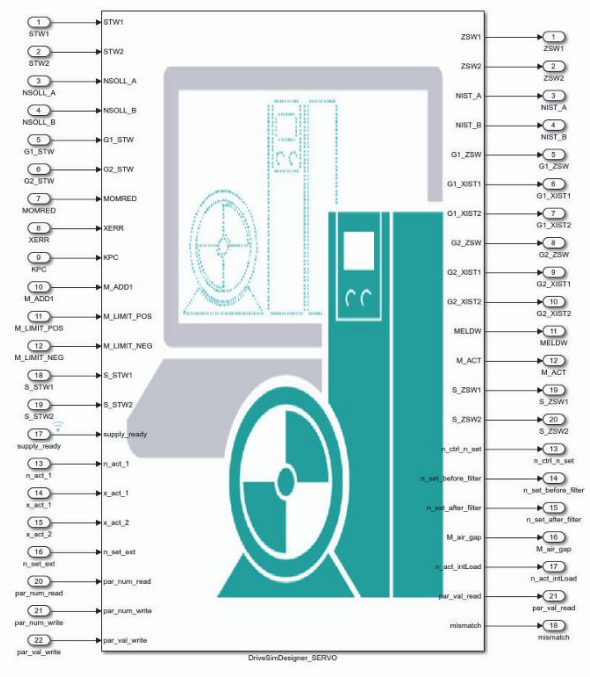


Fig. 1 Drive Sim Designer block in MATLAB

C. Condition monitoring and diagnostics possibilities

Another problem associated with flying shears is the lack of explicit diagnostic information regarding the mechanical state of the cutting mechanism. Standard control systems typically detect only limit violations or faults, while subtle changes in system behavior remain unnoticed.

Digital twin-based analysis enables systematic evaluation of parameters such as speed and torque between expected and measured drive responses. It may indicate increased mechanical stress, misalignment, or changes in cutting conditions.

V. Experimental Setup Using Load Torque Emulator

To enable future validation, an experimental setup consisting of two electric drives is planned. The setup will include an industrial drive system and a real-time control platform based on dSPACE real-time system. The main drive, representing the flying shears, will be controlled by an industrial frequency converter using standard industrial control structures. This ensures realistic operating conditions correspond to practical applications. The second drive will be used as a load torque emulator and will be mechanically coupled to the main drive. Real-time system will be used for supervision, signal acquisition, and the generation of reference signals for the experimental setup. This enables flexible configuration of test scenarios and high-resolution measurement of dynamic responses. By controlling the torque of the load emulator, it will be possible to simulate different operating conditions, including increased cutting resistance and simulated blade wear. This configuration allows realistic and repeatable emulation of load torque without requiring physical degradation of the blades. Signals such as torque, current, and speed will be acquired and compared with outputs of the digital twin. The planned setup will provide a basis for experimental validation of the proposed blade wear identification method. By analyzing these effects over time, the digital twin can support condition monitoring and contribute to early detection of unfavorable operating states.

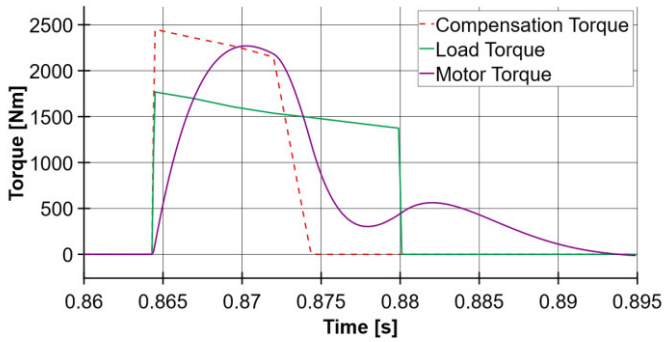


Fig. 2 Illustrative comparison of motor and load torque during cut

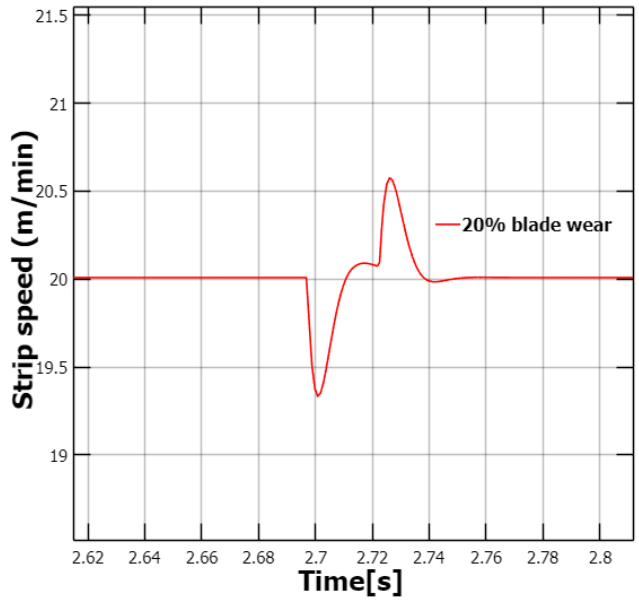


Fig. 3 Speed drop during cutting for 20% blade wear, PI speed control with compensation, AC drive

VI. Conclusion

This paper addressed the possibilities of using a digital twin for flying shears with a focus on the dynamic behavior of the drive system during the cutting phase. It was shown that changes in cutting conditions and the mechanical state of the cutting mechanism, particularly blade wear, have a direct impact on drive load characteristics (Fig. *2*). These effects are reflected in commonly available drive signals such as torque, current, and speed.

By comparing simulated drive responses with measured signals, it is possible to identify changes in dynamic load that are not explicitly evaluated in conventional control systems. Such changes represent valuable indicators of changing operating conditions and gradual degradation of the cutting process.

An important advantage of the proposed approach is that it does not require additional sensors or detailed physical models of blade wear in the initial stages. Instead, blade degradation can be inferred indirectly from its influence on the dynamic response of the drive system. This makes the concept applicable to existing industrial applications and suitable as a basis for condition monitoring.

An important objective of this research will be experimental validation of the proposed approach, which is planned to use a real-time test setup with a torque emulator. This will enable controlled simulation of cutting conditions and verification of the relationship between blade wear and dynamic drive response.

Future work will focus on the evaluation of experimental results and the refinement of the digital twin model, with the aim of developing a method for condition monitoring and predictive maintenance of flying shears systems. The aim will also be to create and define methodology steps for creating digital twin for flying shears, with future potential to extend ecosystem for other parts of material processing lines such as winders, unwinders and others.

Acknowledgement

This work was supported by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00018 and by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the project KEGA 059TUKE-4/2024

References

[1] M. Grieves, *Digital Twin: Manufacturing Excellence through Virtual Factory Replication*, White Paper, 2015. [Online]. Available: http://www.apriso.com/library/Whitepaper_Dr_Grieves_Digital_Twin_ManufacturingExcellence.php

[2] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, “About the importance of autonomy and digital twins for the future of manufacturing,” *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, 2015, doi: 10.1016/j.ifacol.2015.06.141.

[3] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with big data,” *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9–12, pp. 3563–3576, 2018, doi: 10.1007/s00170-017-0233-1.

[4] A. Rasheed, O. San, and T. Kvamsdal, “Digital twin: Values, challenges and enablers from a modeling perspective,” *IEEE Access*, vol. 8, pp. 21980–22012, 2020, doi: 10.1109/ACCESS.2020.2970143.

[5] A. A. Radionov, P. A. Bovshik, B. M. Loginov, A. S. Karandaev, and V. R. Khramshin, “Construction principle for object-oriented digital twins of mechatronic complexes of rolling mills,” in *Proc. Int. Russian Smart Industry Conf. (SmartIndustryCon)*, Sochi, Russia, 2023, pp. 516–522, doi: 10.1109/SmartIndustryCon57312. 2023.10110770.

[6] M. H. Nguyen, M. H. Dao, and J. M. Kim, “Model-based and data-driven approaches for digital twin: A review,” *IEEE Access*, vol. 10, pp. 11612–11636, 2022, doi: 10.1109/ACCESS.2022.3140038.

[7] V. R. Gasiyarov *et al.*, “Substantiating and implementing concept of digital twins for virtual commissioning of industrial mechatronic complexes exemplified by rolling mill coilers,” *Machines*, vol. 11, no. 2, p. 276, 2023, doi: 10.3390/machines11020276.

[8] F. Ďurovský, V. Šlapák, P. Smoleň, K. Kyslan, and E. Spišák, “Rotary shears control in material processing lines,” in *Proc. Int. Conf. Electrical Drives and Power Electronics (EDPE)*, 2023, pp. 1–9, doi: 10.1109/EDPE58625.2023.10274059.

[9] Siemens AG, “DriveSim Designer,” *Siemens Industry Online Support*, Entry ID: 109812859, May 28, 2025. [Online]. Available: https://support.industry.siemens.com/cs/document/109812859

Contribution to modelling and analysis of cyber-physical systems within the framework of the ALICE Detector Control System

[1]*Dominika LÍŠKOVÁ (2nd year),*
Supervisor: [2]*Anna JADLOVSKÁ*

[1,2]Institute of Artificial Intelligence, FEEI, Technical University of Košice, Slovak Republic

[1]dominika.liskova@tuke.sk, [2]anna.jadlovska@tuke.sk

***Abstract*—This paper summarises the tasks undertaken during the past year of the PhD study, focusing on the analysis of the components of the Detector Control System of the ALICE Experiment (ALICE-DCS). It discusses potential modifications of the developed software modules and establishes the requirements for design and implementation of a digital twin of ALICE-DCS. The results of analysis of cyber-physical systems, represented by two DCS architectures described in the paper, provide a foundation for modelling, simulation, testing, and anomaly detection.**

***Keywords*—Cyber-Physical System, Distributed Control System, ALICE Detector Control System, Digital Twin**

I. Introduction

Cyber-physical systems (CPS) represent a class of complex engineered systems in which physical processes are tightly integrated with computation, communication, and control. Their increasing deployment in safety-critical and large-scale applications, such as industrial automation, robotics, smart infrastructures, and scientific experiments—has significantly increased the demands on their modelling, analysis, and validation [1], [2]. Due to their interaction with the physical world and with human operators, CPS are required to operate reliably, safely, and in real time, while coping with uncertainty, communication constraints, and heterogeneous system components [3].

The requirements for implementation of complex distributed systems motivate the implementation of modelling and monitoring approaches, that support structured data archiving, post-operational analysis, and anomaly detection. My dissertation thesis named *Implementation of Hybrid System Models into the Distributed Control Systems Using Modern Methods and Effective Simulation Tools* is solved within the basic research project entitled *ALICE Experiment at the LHC at CERN: Study of Strongly Interacting Matter in Extreme Conditions (ALICE-CERN at TUKE)*, with the selected tasks reflecting the partial objectives of the project. The research objectives are conducted at the ALICE Experiment at CERN, as well as at the home institution within the Centre of Modern Control Techniques and Industrial Informatics (CMCT&II) at the Technical University of Košice (TUKE).

II. Initial stage of research

The increasing complexity of cyber-physical systems, particularly in safety-critical and distributed environments, raises significant challenges in their modelling, analysis, and validation. In such systems, the tight integration of physical processes, computation, and communication requires the use of methods capable of representing both discrete control logic and continuous dynamics.

We initially focused on analysing existing modelling approaches for CPS, with an emphasis on distributed and hierarchical architectures. Various representations of hybrid systems were reviewed; with a focus on finite-state machines and hybrid automata [4]. In parallel, distributed control systems were examined as the dominant architecture for CPS, characterised by hierarchical, networked structures formalised by standards such as IEC 61499 and IEC 62264 [5], [6].

The mentioned project ALICE-CERN at TUKE focuses on the ALICE Experiment at CERN. ALICE (A Large Ion Collider Experiment) is one of the four major experiments at the Large Hadron Collider (LHC), dedicated to the study of strongly interacting matter under extreme conditions. The experiment relies on a complex and distributed control infrastructure. During LHC collisions, the ALICE experiment acquires large volumes of both physics and process data. The process data are monitored and processed by the ALICE Detector Control System (ALICE-DCS), which ensures the stable and safe operation of the detectors and subsystems of the experiment [7].

In order to describe the practical implementation of DCS, two architectures were analysed: the ALICE-DCS at CERN and the DCS architecture at CMCT&II at TUKE, both organised as five-level architectures with structural similarities.

The ALICE-DCS architecture comprises these five levels: the level of detectors, sensors, and actuators (level 0); the level of technological control of sub-detectors and front-end devices (level 1); the SCADA/HMI level (level 2), including visualisation and control managed by WinCC; the information level (level 3) represented by databases storing data from lower-level electronics and sub-detectors; and the management level (level 4) consisting of a worldwide network of GRID computing centres and archival databases [7].

The five-level DCS architecture at CMCT&II at TUKE is

designed similarly to the ALICE-DCS architecture, with slight differences in the individual levels. These levels include: the level of models, sensors, and actuators (level 0); the level of technological control and regulation (level 1); the level of simulation models and SCADA/HMI (level 2); the information level (level 3); and the management level (level 4). This DCS architecture represents a CPS organised into hierarchical levels, enabling modularity and structured integration of physical and software components. [8].

To further describe the challenges and requirements of CPS, the communication and control infrastructure of the ALICE-DCS has been analysed. As shown in Fig. 1, the first three levels of the ALICE-DCS architecture correspond to the frontend electronics level (level 0), the middleware level called device access and hardware abstraction level (level 1), and the supervisory control level (level 2), specified as central systems, implemented in WinCC OA. The middleware level (level 1) is implemented through the ALFRED framework, which provides abstraction between hardware and SCADA and ensures coordinated data exchange and command execution across the distributed infrastructure.

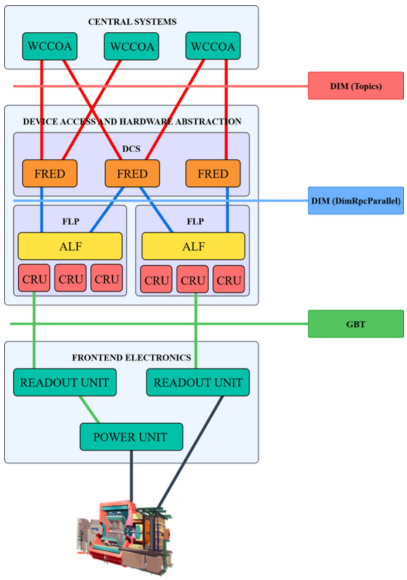


Fig. 1. Communication and control infrastructure of ALICE-DCS

The ALFRED framework has been developed within three stages of the ongoing project ALICE-CERN at TUKE. The ALFRED framework consists of the ALF and FRED modules, abstracts low-level hardware communication, enables modular and parallel control of frontend electronics, and maps hardware registers to standardised process variables [9]. By relocating sequencing, control logic and data processing from the SCADA level to ALFRED framework, the architecture ensures scalability, transparency, and service-oriented access for detector experts without requiring modifications to the central DCS infrastructure [10].

III. Solved tasks

The following tasks were addressed as part of the future dissertation thesis. Firstly, we focused on the analysis of the existing modules and exploring potential modifications to accommodate evolving operational conditions within the ALICE-DCS. During mobility and a summer internship at CERN, the solved tasks involved direct interaction with the system, evaluation of module behaviour, and identification of areas for improvement or adaptation.

The analysis of ALICE-DCS components included a review of specific software modules. Our attention was directed to the first level shown in Fig. 1, which includes the abstraction layer and the ALFRED framework. Hybrid system models were applied to represent module behaviour: finite-state machines (FSM) were used for the ALF and FRED modules, while Petri nets captured communication and interactions across levels. This combination of modelling approaches enabled the simulation of individual module behaviour as well as inter-module interactions within the operational environment [10]. The results of this analysis provided insight into the implementation of existing modelling approaches within the ALICE-DCS architecture and identified areas suitable for modifications and improvements, reflecting the highly dynamic and evolving nature of the ALICE experiment.

Each detector module in the ALICE experiment is unique and requires specific readout chains, control systems, and supporting infrastructure. Most subsystems are prototypes without reference models or digital twins, and laboratory setups often include only partial modules or substitute hardware, limiting testing and increasing the risk of damage to sensitive electronics.

To overcome these limitations, a digital twin (DT) of the ALICE-DCS will be developed to model individual components and their interconnections. To enable the development of the DT of the ALICE-DCS, we designed a model application, illustrated in Fig. 2, based on the analysis and implementation of the ALICE-DCS components. The model application (Fig. 2) schematically shows the modules of ALICE-DCS. It extends the three lowest levels shown in Fig. 1, with the third level represented by an Oracle Database and the fourth level comprising the software modules for data retrieval. In parallel with the lowest level, which represents the detector electronics and ALF module, the model application includes the ALF-SIM software module, which was implemented within the project ALICE-CERN at TUKE. The purpose of the ALF-SIM module is to emulate the front-end subsystems, and to replicate the communication with FRED, including the outputs compatible with the ALICE-DCS.

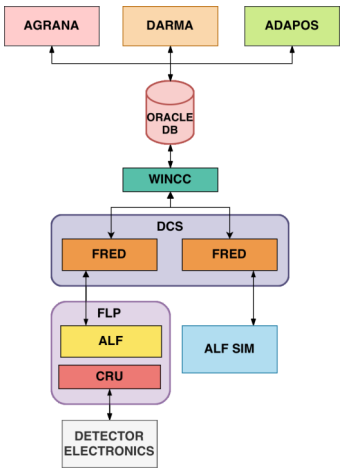


Fig. 2. Representative application for ALICE-DCS

The ALF-SIM software module integrates with existing workflows, allowing development and testing without physical hardware. Simplified hardware devices, such as IoT-based

temperature sensors, can also be used to provide equivalent measurements for hybrid system inputs, enabling controlled testing of error scenarios. At the supervisory level, a SCADA simulation module can replace the physical WinCC OA system by generating realistic data points, performing monitoring, archiving, and alarm generation.

As shown in Fig. 2, data from the detector devices pass through the ALFRED framework to the supervisory level represented by WinCC OA, and are subsequently archived in an Oracle database. To access and process these detector data, the software modules AGRANA, DARMA, and Adapos have been introduced.

Within the scope of the ALICE-CERN at TUKE project, we contributed to the development and implementation of the AGRANA and DARMA modules. DARMA supports offline analysis, allowing detector experts to query the archival database and study detector behaviour over time. AGRANA provides online, real-time monitoring through interactive dashboards, enabling operators to track the current state of the ALICE experiment and analyse critical parameters. These tools ensure both detailed analysis and immediate situational awareness, supporting the safe and efficient operation of the experiment. The results of the development and application of these modules have been published in the authors' paper [11].

As a result of the summer internship, the portfolio of hardware and software requirements was analysed and prepared, providing the basis for the implementation and validation of the developed modules. The prerequisites for developing a digital twin were established by combining physical devices, software modules, and simulation models across levels, allowing the full control chain from detectors through ALF, FRED, and SCADA systems to be represented. This approach supports software development, testing, and validation, illustrating the potential for a practical implementation of a digital twin for ALICE-DCS. The AGRANA and DARMA modules will also be used to access and analyse process data as part of the future dissertation thesis, forming a link between data acquisition, system modelling, and data analysis, including anomaly detection.

CPS principles were applied to the design and implementation of a model application for a mobile robotic platform. The objective was to create a scalable and modular mobile robotic platform enabling integration of heterogeneous sensors and actuators, and testing of control strategies. The CPS architecture of the robotic platform was directly inspired by the DCS architecture defined at CMCT&II and described in Section II, adopting its hierarchical structure with slight modifications in the level nomenclature: component (level 0), control (level 1), coordination (level 2), supervisory (level 3), and management (level 4).

This approach also enabled the implementation of a digital twin of the robotic platform, supporting simulation of component interactions, testing of localisation algorithms, and analysis of operational scenarios. In addition to the system design and implementation, a significant outcome of this work was the authors' publication of the achieved results in [6].

IV. Future research

Future research will focus on modification of the models and simulation environment to enable comprehensive representation and validation of ALICE-DCS components, further development of the AGRANA and DARMA modules, and the establishment of requirements for a digital twin. The planned extensions include the integration of real and simulated data streams across multiple levels, the evaluation of the system stability and robustness, predictive maintenance and anomaly detection. The digital twin will provide a vertical view of the entire experiment and a flexible, controlled environment for tasks such as system benchmarking, throughput testing, and incident simulations, while enabling systematic monitoring of system behaviour and inter-component relationships.

V. Conclusion

This paper presents an analysis of distributed control systems, focusing on the components of the ALICE-DCS at CERN and the DCS architecture at CMCT&II at TUKE. It outlines the requirements for implementing digital twins for both architectures and summarises the developed modules that provide a foundation for their creation. Together, these results establish the basis for further modelling, simulation, and advanced data analysis with a focus on anomaly detection.

Acknowledgement

This work has been supported by the project ALICE experiment at the CERN LHC: The study of strongly interacting matter under extreme conditions (ALICE TUKE 0410 / 2022).

References

[1] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proceedings of the 47th Design Automation Conference*, ser. DAC '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 731–736. [Online]. Available: https://doi.org/10.1145/1837274.1837461

[2] Y. Eslami, C. Franciosi, S. Ashouri, and M. Lezoche, "A review and analysis of the characteristics of cyber-physical systems in industry 4.0," *SN Computer Science*, vol. 4, no. 6, p. 825, 2023. [Online]. Available: https://doi.org/10.1007/s42979-023-02268-0

[3] X. Guan, B. Yang, C. Chen, W. Dai, and Y. Wang, "A comprehensive overview of cyber-physical systems: from perspective of feedback system," *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 1, pp. 1–14, 2016. [Online]. Available: https://doi.org/10.1109/JAS.2016.7373757

[4] D. Líšková, "Modelling and analysis of cyber-physical systems," in *25th Scientific Conference of Young Researchers (SCYR 2025) – Proceedings*. Košice, Slovakia: Technical University of Košice, 2025, pp. 52–55.

[5] E. Monroy Cruz, L. R. García Carrillo, and L. A. Cruz Salazar, "Structuring cyber-physical systems for distributed control with iec 61499 standard," *IEEE Latin America Transactions*, vol. 21, no. 2, pp. 251–259, 2023. [Online]. Available: https://doi.org/10.1109/TLA.2023.10015217

[6] D. Líšková, A. Jadlovská, and F. Pazdič, "A CPS-Based Architecture for Mobile Robotics: Design, Integration, and Localisation Experiments," *Sensors*, vol. 25, no. 18, p. 5715, Sep. 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/18/5715

[7] P. Chochula, A. Augustinus, M. P. Bond, N. A. Kurepin, M. Lechman, D. Vošček, and O. Pinazza, "15 years of alice dcs," in *Proceedings of ICALEPCS 2023*, Cape Town, South Africa, 2023, pp. 1002–1007. [Online]. Available: https://jacow.org/icalepcs2023/papers/we2bco07.pdf

[8] A. Jadlovská, S. Jadlovská, and D. Vošček, "Cyber-physical system implementation into the distributed control system," *IFAC-PapersOnLine*, vol. 49, no. 25, pp. 31–36, 2016, 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016. [Online]. Available: https://doi.org/10.1016/j.ifacol.2016.12.006

[9] M. Tkáčik, J. Jadlovský, S. Jadlovská, L. Koska, A. Jadlovská, and M. Donadoni, "Fred—flexible framework for frontend electronics control in alice experiment at cern," *Processes*, vol. 8, no. 5, 2020.

[10] M. Tkáčik, J. Jadlovský, S. Jadlovská, A. Jadlovská, and T. Tkáčik, "Modeling and analysis of distributed control systems: Proposal of a methodology," *Processes*, vol. 12, no. 1, 2024.

[11] D. Líšková, A. Jadlovská, S. Jadlovská, P. Chochula, and P. Bond, "Online data access and visualisation interfaces of the detector control system within the alice experiment at cern," *Acta Electrotechnica et Informatica*, 2026, accepted for publication.

Analysis, Identification, Modeling and AI-Based Compensation of Cogging Torque

[1]*Tadeáš KMECIK (3rd year)*
Supervisor: [2]Peter GIROVSKÝ

[1,2] Dept. of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]tadeas.kmecik@tuke.sk, [2]peter.girovsky@tuke.sk

***Abstract*— The publication investigates cogging torque in PMSM and BLDC motors and reviews both design- and control-based suppression methods. It proposes a simple identification methodology implemented in a MATLAB/Simulink simulation model, based on the standard Field Oriented Control (FOC) structure. The approach enables construction of a cogging torque map as a function of rotor position without modifying control loop parameters, and this map is used to generate a compensation signal in the torque-producing current reference. In addition, a novel method for determining an ideal compensation map using a predictive neural model of the torque control loop is introduced. The neural-based approach, which accounts for drive dynamics and loop interactions, is compared with conventional map-based compensation. The effectiveness of both methods is assessed through the reduction of angular velocity ripple.**

***Keywords*—PMSM, BLDC, compensation, identification, FOC, neural model, MATLAB/Simulink**

I. INTRODUCTION

Permanent magnet motors are among the most widely used electric drives due to their high-power density, efficiency, precise controllability, and reliable operation. They are applied across a broad range of applications, from household appliances to electromobility and robotics [1].

However, the pressure to reduce manufacturing costs leads to simplified designs, which may increase the occurrence of parasitic torques, particularly cogging torque, as well as frictional and other mechanical or electromagnetic effects. These phenomena negatively affect smooth operation [1].

The reduction of parasitic torques can be achieved through either design modifications or control-based interventions. Design solutions, such as stator skewing, magnet skewing, or magnet shape optimization, are effective but increase manufacturing complexity and cost [2].

An alternative to structural modifications is the compensation of cogging torque using control algorithms that inject a compensation signal into the control structure. The advantage of this approach is the possibility of application to existing motors without mechanical modifications. The disadvantages include higher computational requirements and a potential slight reduction in efficiency.

Compensation is particularly challenging due to the high frequency of cogging torque relative to the stator frequency. Standard PI controllers within the FOC structure have limited bandwidth, which results in phase delay and attenuation of higher harmonics. Compensation is therefore most problematic at medium and higher speeds, although the effect of cogging torque is partially damped there by the drive inertia [2][3].

Cogging torque is most pronounced at low speeds, where the drive inertia is unable to effectively suppress its effect. It manifests as angular velocity ripple, increased vibrations, and acoustic noise, which may lead to higher mechanical stress in the connected system and reduced user comfort, for example in household appliances [3][4].

Advanced approaches reported in the literature include high-bandwidth controllers, model-based torque observers, and neural networks; however, their implementation in low-cost drives is limited by the computational capabilities of the control hardware [3][5][6].

Works [7]-[9] address cogging torque reduction through structural modifications, which are effective but cannot be applied to already manufactured motors. Control-based approaches include, for example, the PIR (proportional–integral–resonant) controller [3], which suppresses harmonic components using resonant terms, albeit at the cost of increased implementation and computational complexity. The IMR (internal model resonant) controller [10] enables adaptive compensation of higher harmonics but requires online harmonic analysis and higher computational performance. Model-based approaches estimate cogging torque parameters in real time; however, their effectiveness depends on model accuracy and proper tuning. These methods are computationally demanding.

The cogging torque map can be obtained using direct or indirect methods. Direct measurement employs a dedicated torque sensor, providing high accuracy but at increased cost [7][8]. Such a solution is economically justified mainly in mass production, where the map can be supplied directly by the motor manufacturer.

Indirect methods rely on available position and current sensors [9]. In this work, an indirect approach based on evaluating the torque-producing current component is verified. Since this component is proportional to electromagnetic torque, it enables identification of the cogging torque map.

The current waveform over one mechanical revolution is processed using the Fast Fourier Transform (FFT) to identify dominant harmonic components, including their amplitudes and phases. Based on these components, the cogging torque is modeled in MATLAB.

The structure of the paper is as follows: The second section focuses on the analysis, identification, and modeling of cogging torque. The third section presents a novel AI-based approach for obtaining an optimal compensation signal.

The conclusion summarizes the achieved results, identifies the limitations of the proposed approaches, and outlines possible directions for further research in cogging torque compensation.

II. Analysis, Modeling and Identification of Cogging Torque

A. Physical and Harmonic Analysis of Cogging Torque

Cogging torque is a parasitic phenomenon arising from the interaction between the magnetic field of the rotor permanent magnets and the ferromagnetic stator slots.

From a physical standpoint, cogging torque can be expressed by the following relation:

$$T_{COGG} = -\frac{\Phi_{PM}{}^2}{2}\frac{dR(\vartheta_r)}{d\vartheta_r} \quad (1)$$

where $R(\vartheta_r)$ represents the reluctance of the magnetic circuit as a function of the rotor angular position ϑ_r and Φ_{PM} denotes the magnetic flux generated by the permanent magnets.

In addition to the physical formulation, cogging torque can also be expressed using a Fourier series, which enables its efficient analysis and synthesis:

$$T_{COGG} = \sum A_{nT}\cos(nLCM\vartheta_r + \phi_n) \quad (2)$$

where A_{nT} is the amplitude of the n-th harmonic component, n is its order, denotes the least common multiple of the number of stator slots and rotor poles, and ϕ_n is the phase of the respective component. The LCM determines the fundamental harmonic of the cogging torque, which is typically dominant [3].

FFT is used to identify dominant harmonics for generating the compensation signal. In practice, the waveform may also contain components lower than the LCM multiple [3]. The simulated motor has 12 stator slots and 6 rotor poles (LCM = 12).

B. Modeling of Cogging Torque

In the simulation, the cogging torque map was modeled based on known characteristics of its amplitude and frequency spectrum.

The amplitude was selected at approximately 10% of the rated torque, corresponding to typical values reported in the literature for low-cost motors. The model includes not only the fundamental harmonic component but also selected lower-order harmonics (half-order and third-order) to more realistically represent the actual waveform. The resulting cogging torque was implemented in the simulation as a load torque acting on the motor [11]. The waveform of the modeled cogging torque is shown in *Fig. 1*.

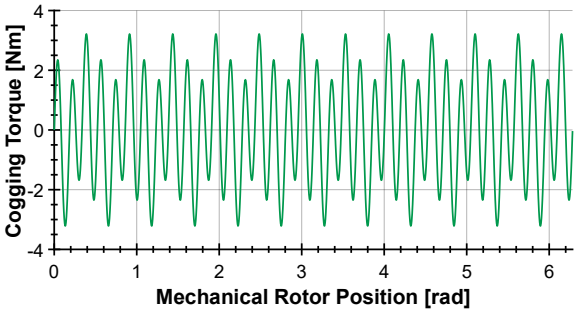


Fig. 1 Cogging torque profile as a function of rotor position

C. Identification of the Cogging Torque Map

The cogging torque map was identified from the waveform of the torque-producing current component, which is proportional to the electromagnetic torque under constant flux conditions. The identification was performed at a very low speed in order to minimize the influence of controller dynamics, while the sampling of the q-axis current was carried out as a function of rotor position.

The acquired data (q-current and position) were processed using FFT analysis, from which the amplitudes and phases of the dominant harmonic components *Tab. 1* were determined. Based on these components, a reference compensation signal was subsequently constructed. The objective was to verify the possibility of determining the cogging torque waveform solely from current and position measurements, without using a torque sensor. A comparison with the modeled waveform confirmed the validity of the method, with only a slight phase shift observed due to control loop dynamics.

Tab. 1 Result of FFT analysis

n-harmonic	A_{nI}	ϕ_n
1(DC-component)	0/k_t	-
12	3/k_t	(3/2)π
24	1/k_t	-π/2
36	0.5/k_t	π/2

D. Cogging Torque Compensation Modeling

The reference waveform was derived from the FFT results presented in *Tab. 1* and represents the modeled cogging torque. Based on this waveform, a compensation current was constructed according to equation (3), which is added to the reference value of the torque-producing current component.

$$I_{qcomp} = \sum A_{nI}\cos(n\vartheta_r + \phi_n) \quad (3)$$

Within this publication, the reference compensation signal will be compared with an ideal compensation waveform obtained using a neural model of the motor torque control loop. The objective of this comparison is to verify to what extent a predictive neural network (NN) model can achieve more accurate cogging torque compensation.

III. Neural Model of the Torque Loop for Optimal Cogging Torque Compensation

The neural model of the torque control loop is designed as a dynamic approximation model implemented using a multilayer perceptron (MLP) NN. The MLP consists of two hidden layers, each containing eight neurons. The inputs to the model consist of the reference torque, the estimated torque, and their one-sample delayed values, enabling the capture of the system dynamics. The model output represents the predicted motor torque in the next discrete time step. The model is trained on system data to reproduce the behavior of the real torque control loop as accurately as possible.

After identification, the model is used as a predictive tool to determine the ideal compensation signal. The fundamental idea is that if the motor generates a torque equal in magnitude to the cogging torque (with opposite sign), the resulting torque acting on the shaft will be zero, and thus no change in angular velocity will occur.

For each rotor position, the algorithm therefore searches for such a torque reference value that, according to the neural

model, ensures a zero resultant torque.

Since the cogging torque is uniquely dependent on rotor position, this procedure makes it possible to identify the ideal compensation torque waveform for a given operating speed.

IV. Simulation Results

Based on the identified cogging torque map derived from the torque-producing current component, a reference compensation signal was constructed using the dominant harmonic components listed in *Tab. 1* The signal was added to the output of the speed PI controller, which defines the reference torque (i.e., the reference current).

This static map-based compensation was compared with a neural network NN-optimized approach. Unlike the conventional method, which directly uses the identified harmonics, the NN-based compensation accounts for torque loop dynamics and determines, for each rotor position, the torque reference that minimizes the difference between the generated torque and the cogging torque, thereby reducing angular velocity ripple.

The performance of both methods is compared in *Fig. 2, Fig. 3, Fig. 4.*

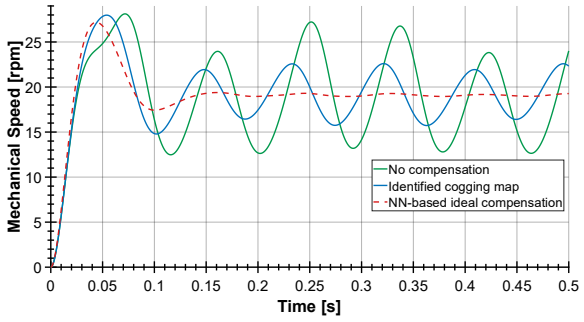


Fig. 2 Angular velocity ripple for three compensation states at 2 rad/s.

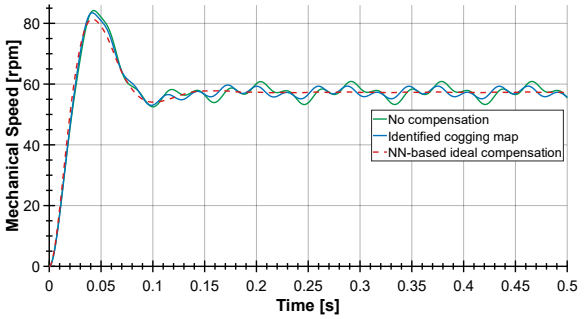


Fig. 3 Angular velocity ripple for three compensation states at 6 rad/s.

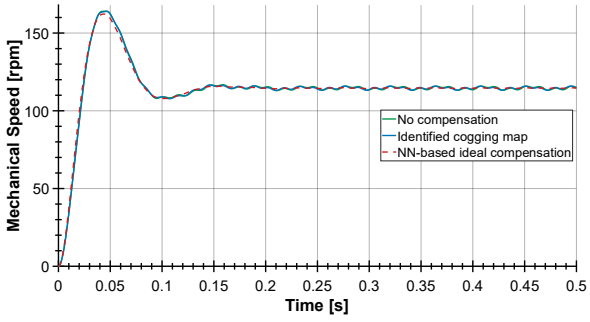


Fig. 4 Angular velocity ripple for three compensation states at 12 rad/s.

V. Conclusion

Cogging torque compensation is most critical at low speeds, where it significantly reduces angular velocity ripple. At higher speeds, the effect of cogging torque is partially attenuated by motor inertia, and the benefit of compensation decreases.

The static map-based compensation is most effective near zero speed; however, its performance deteriorates with increasing speed due to phase shift and amplitude mismatch. In contrast, the NN-optimized compensation achieves nearly negligible speed ripple across the tested operating range.

These results validate the proposed method and demonstrate its capability to identify an ideal compensation map over the full operating range. The resulting waveform can be stored as a lookup table (LUT) in the control microcontroller.

The main contribution lies in achieving effective compensation with low computational demands, enabling implementation on standard embedded hardware and supporting the use of simpler, lower-cost PMSM and BLDC motors with software-based cogging torque suppression.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under Contracts No. APVV-19-0210 and APVV-23-0521, and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under Project KEGA 059TUKE-4/2024.

References

[1] SORESINI, Federico, et al. Noise and Vibration Modeling of Permanent Magnet Synchronous Motors: A Review. IEEE Transactions on Transportation Electrification, 2024, 10.4: 8728-8745.

[2] N. Bianchi and S. Bolognani, "Design techniques for reducing the cogging torque in surface-mounted PM motors," in IEEE Transactions on Industry Applications, vol. 38, no. 5, pp. 1259-1265, Sept.-Oct. 2002, doi: 10.1109/TIA.2002.802989.

[3] C. Xia, B. Ji and Y. Yan, "Smooth Speed Control for Low-Speed High-Torque Permanent-Magnet Synchronous Motor Using Proportional–Integral–Resonant Controller," in IEEE Transactions on Industrial Electronics, vol. 62, no. 4, pp. 2123-2134, April 2015, doi: 10.1109/TIE.2014.2354593.

[4] KRISHNAN, Ramu. Permanent magnet synchronous and brushless DC motor drives. CRC press, 2017.

[5] J. Liu, H. Li and Y. Deng, "Torque Ripple Minimization of PMSM Based on Robust ILC Via Adaptive Sliding Mode Control," in IEEE Transactions on Power Electronics, vol. 33, no. 4, pp. 3655-3671, April 2018, doi: 10.1109/TPEL.2017.2711098.

[6] A. Mora, Á. Orellana, J. Juliet and R. Cárdenas, "Model Predictive Torque Control for Torque Ripple Compensation in Variable-Speed PMSMs," in IEEE Transactions on Industrial Electronics, vol. 63, no. 7, pp. 4584-4592, July 2016, doi: 10.1109/TIE.2016.2536586.

[7] G. Dajaku and D. Gerling, "New methods for reducing the cogging torque and torque ripples of PMSM," 2014 4th International Electric Drives Production Conference (EDPC), Nuremberg, Germany, 2014, pp. 1-7, doi: 10.1109/EDPC.2014.6984396.

[8] M. Y. Arafat, M. Murshed, M. M. Hasan and M. A. Razzak, "Impacts of cogging torque and its reduction for an external rotor permanent magnet alternator," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2016, pp. 536-541, doi: 10.1109/ICIEV.2016.7760061.

[9] S. Leitner, H. Gruebler and A. Muetze, "Low-Cost BLDC Claw-Pole Motor Design for Fan Applications with Reduced Cogging Torque and Balanced Axial Forces," 2020 IEEE Applied Power Electronics Conference and Exposition (APEC), New Orleans, LA, USA, 2020, pp. 279-284, doi: 10.1109/APEC39645.2020.9124114.

[10] E. H. Abd-Elhameed and M. Iwasaki, "Pulsating torque modeling and compensation for high precision positioning control," 2009 IEEE International Conference on Mechatronics, Malaga, Spain, 2009, pp. 1-5, doi: 10.1109/ICMECH.2009.4957142.

[11] F. Bu et al., "Speed Ripple Reduction of Direct-Drive PMSM Servo System at Low-Speed Operation Using Virtual Cogging Torque Control Method," in IEEE Transactions on Industrial Electronics, vol. 68, no. 1, pp. 160-174, Jan. 2021, doi: 10.1109/TIE.2019.2962400.

Progress in experimental investigation of discharges in nanofunctionalized transformer oil

[1]*Kristián GLAJC (2nd year)*
Supervisor: [2]Juraj KURISMKÝ

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]kristian.glajc@tuke.sk, [2]juraj.kurimsky@tuke.sk

***Abstract*— The reliability of power transformers depends on the behaviour of liquid insulation under impulse stress. This paper presents the development and experimental validation of a measuring platform designed for impulse discharge investigation in transformer oil. Preliminary measurements were performed in Diala S5 under positive impulse polarity up to 220 kV (1.3 µs) to verify the dielectric withstand capability of the constructed test cell and the stability of the impulse configuration. Interchangeable protective resistances were implemented to enable controlled regulation of discharge energy and current. The experiments confirmed reproducible discharge initiation and demonstrated the suitability of the platform for optical documentation.**

***Keywords*— impulse voltage; transformer oil; streamer propagation; nanofluids.**

I. INTRODUCTION

Reliable operation of power transformers is critically dependent on the condition and performance of their insulation systems. Transformer oil serves both as an electrical insulating medium and as a cooling agent, and its dielectric strength directly influences the operational safety of high-voltage equipment. Under impulse overvoltage conditions, electrical breakdown in transformer oil is typically preceded by the formation and propagation of streamers, which play a decisive role in determining breakdown voltage and discharge development [1], [2].

The physics of streamer propagation in liquid dielectrics has been investigated using numerical electrodynamic models based on Gauss's law and charge continuity equations for electrons, ions, and charged nanoparticles. Such models describe charge generation through field ionization, drift-dominated particle transport, recombination processes, and the feedback effect of space charge on the local electric field distribution. Numerical simulations enable detailed analysis of ionization wave dynamics and streamer development in both pure and nanofunctionalized transformer oils [3],[4]. Although these models provide deep physical insight into discharge mechanisms, experimental verification remains essential for accurate interpretation of real discharge phenomena.

In recent years, increasing attention has been devoted to nanofunctionalized transformer oils, where dispersed nanoparticles are introduced to enhance dielectric properties. Experimental studies indicate that nanoparticle type, concentration, and dispersion quality may significantly influence charge transport processes, local electric field distribution, and streamer dynamics [5], [6].

Various diagnostic techniques have been employed for investigation of discharges in liquid insulation systems, including electrical measurements, acoustic emission methods, and optical visualization [1],[2]. Among these approaches, high-speed optical diagnostics provides direct information about streamer morphology, propagation velocity, and branching behaviour, which cannot be fully resolved using purely electrical methods.

The present research therefore focuses on the experimental investigation of streamer discharges in pure and nanofunctionalized transformer oil under impulse voltage stress, with emphasis on controlled optical observation of pre-breakdown streamer development.

II. RESEARCH PROGRESS DURING THE ACADEMIC YEAR

During the evaluated academic year, research activities were focused on establishing a suitable experimental setup for impulse discharge investigations in transformer oil.

The experiments were conducted in the High Voltage Technology Laboratory at the Faculty of Electrical Engineering and Communication (FEKT), Brno University of Technology (VUT Brno). The laboratory is equipped with a comprehensive high-voltage impulse testing system for controlled generation and shaping overvoltage stress conditions, as shown in Fig. I.

Haefely SG 60-1000/100 Impulse Voltage Generator: This generator can produce standardized high-voltage impulses with peak amplitudes up to 1000 kV. It serves as the primary source of controlled overvoltage stress for liquid dielectric breakdown testing and supports a range of impulse waveforms in accordance with IEC standards.

Haefely MAFS 1200/1200 Multiple Chopping Gap: The chopping gap unit enables shaping of the impulse front and tail. By interrupting the impulse discharge at a predefined level, the MAFS system refines the rise time and waveform characteristics, providing a repeatable and adjustable impulse shape essential for consistent testing and comparison.

Haefely CS 1000-670 Impulse Driver Unit: This driver module provides precise triggering, timing control, and synchronization of the impulse generator with the measurement and diagnostic systems. It ensures that voltage application, data acquisition, and optical

diagnostic equipment operate in a coordinated manner, which is crucial for repeatable and synchronized measurements.

This configuration allows controlled generation and shaping of lightning- and switching-type impulse voltages, making it suitable for detailed investigations of discharge initiation, propagation, and dielectric response under high-voltage stress.

Fig. I High-voltage impulse testing system installed in the High Voltage Technology Laboratory (FEKT, VUT Brno): (1) Haefely SG 60-1000/100 impulse voltage generator, (2) Haefely MAFS 1200/1200 Multiple Chopping Gap, (3) Haefely CS 1000-670 impulse driver unit.

A dedicated test cell was designed and constructed to enable controlled impulse voltage experiments. The cell was manufactured from insulating material and equipped with a needle–plane electrode configuration with a fixed 20 mm gap. Optical access was provided by a plexiglass window sealed with silicone to ensure oil tightness. Threaded ports were implemented for oil filling and venting, allowing complete filling of the measuring volume without trapped air.

The impulse voltage circuit was configured with interchangeable protective resistors (6 kΩ, 50 kΩ, and 300–500 kΩ) to regulate discharge current and prevent uncontrolled energetic breakdown during preliminary validation tests.

Preliminary impulse measurements were carried out using pure transformer oil (Diala S5). The primary objective was to verify the dielectric withstand capability of the constructed cell and to evaluate its suitability for controlled discharge initiation and optical documentation.

For optical documentation, a digital camera system Nikon, was employed for preliminary visualization of discharge events. High-speed optical diagnostics will be implemented in subsequent experimental phases.

III. Result And Discussion

Impulse voltage tests were performed under positive polarity with a front time of approximately 1.3 µs and peak voltages ranging from 140 kV to 220 kV (Table 1.).

The constructed test cell withstood impulse voltages up to 220 kV without mechanical failure, confirming its suitability for high-voltage experiments.

TABLE I
Applied impulse voltage levels and corresponding discharge outcomes

Impulse voltage [kV]	Resistance [kΩ]	Streamer Observed
120	6	-
130	6	-
135	6	✓
140	6	✓
149	6	✓
159	50	✓
160	50	✓
200	300-500	-
220	300-500	✓
220	300-500	✓

A significant influence of protective resistance on discharge development was observed. At low resistance (6 kΩ), rapid transition from streamer inception to a fully developed conductive channel occurred. The discharge bridged the entire 20 mm gap, forming a highly luminous plasma channel (Fig. II).

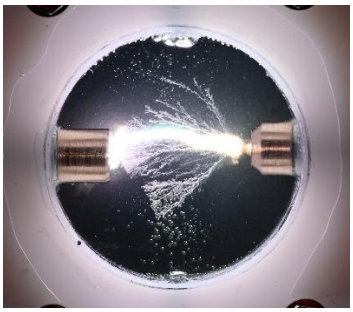


Fig. II Discharge development under positive impulse voltage 149 kV with 6 kΩ protective resistance.

At intermediate resistance (50 kΩ), full breakdown across the electrode gap was still observed; however, the discharge development exhibited lower apparent energy compared to the 6 kΩ configuration.

For higher resistance values (300–500 kΩ), discharge development appeared more controlled, although complete gap bridging remained possible. These observations qualitatively confirm that increasing protective resistance limits discharge current and reduces the energy delivered to the discharge channel.

Impulse testing under negative polarity resulted in external flashover outside the test cell, indicating the need for further optimization of the feeding configuration.

Gas bubble formation was observed during impulse testing and is assumed to be related to localized heating effects or incomplete degassing of the oil.

Optical documentation using a conventional digital camera was performed solely for preliminary verification of discharge occurrence. Detailed analysis of pre-breakdown streamer dynamics will be carried out using high-speed optical diagnostics in subsequent experiments.

During the evaluated academic year, the impulse test cell was designed and constructed, the protective resistance control

system was implemented, validation measurements up to 220 kV were performed, eight nanofunctionalized oil samples with defined concentrations were prepared, and experimental cooperation with the High Voltage Technology Laboratory at Brno University of Technology was established.

In parallel with the experimental activities, the student prepared a review paper focused on a systematic comparison of methods used for investigation of electrical discharges in nanofunctionalized transformer oils, including electrical, optical, acoustic, and hybrid diagnostic approaches. The manuscript is currently under preparation for submission to a scientific journal.

These activities resulted in the successful development and establishment of a solid theoretical and methodological basis for subsequent detailed investigation of discharge mechanisms in nanofunctionalized transformer oils.

IV. Future works

Future research will be conducted at the Faculty of Electrical Engineering and Communication (FEKT), Brno University of Technology (VUT Brno), where advanced high-speed optical diagnostics are available.

The primary objective will be detailed investigation of the pre-breakdown phase of streamer development prior to the formation of a stable conductive channel. Emphasis will be placed on quantifying streamer propagation velocity and morphological characteristics in pure transformer oil and in nanofunctionalized oils containing C60 nanoparticles with concentrations of 0.005 wt.%, 0.01 wt.%, 0.02 wt.%, and 0.03 wt.% and magnetic nanoparticles with concentrations of 0.01 wt.%, 0.02 wt.%, 0.05 wt.%, and 0.1 wt.%.

In addition to optical diagnostics, systematic measurements of partial discharge inception voltage (PDIV) and extinction voltage (PDEV) will be performed to evaluate the influence of nanoparticle type and concentration on discharge initiation and stability.

Further optimization of the experimental setup will include improvement of optical access conditions, implementation of temperature monitoring, enhancement of degassing procedures, and refinement of the impulse feeding configuration to ensure stable and reproducible measurement conditions.

These activities will enable comprehensive investigation of discharge mechanisms in nanofunctionalized transformer oils through combined electrical and optical diagnostic approaches.

V. Conclusion

The preliminary impulse experiments verified the mechanical integrity and dielectric withstand capability of the constructed test cell under impulse voltages up to 220 kV. The applied configuration enabled reproducible discharge initiation under positive polarity in a 20 mm needle–plane arrangement, confirming the suitability of the platform for controlled impulse testing.

The influence of series protective resistance on discharge behaviors was qualitatively confirmed. Adjustment of resistance values allowed regulation of discharge current and energy input into the discharge channel, demonstrating the capability of the setup to control discharge development conditions.

In parallel with the experimental activities, a systematic review of diagnostic techniques for investigation of electrical discharges in liquid dielectrics was conducted. The analysis covered electrical, optical, acoustic, and hybrid approaches and emphasized the importance of integrating electrical measurements with optical diagnostics for comprehensive characterization of discharge phenomena.

Overall, the experimental validation and methodological analysis established a reliable research platform for subsequent detailed investigation of discharge mechanisms in nanofunctionalized transformer oils. The developed setup provides a technical foundation for future high-speed optical diagnostics and comparative nanoparticle studies.

Acknowledgment

This research was funded by the Ministry of Education, Research, Development and Youth within the project VEGA 2/0029/24, 1/0380/24, the Slovak Agency for Research and Development based on contracts no. APVV-22-0115.

The work also received support from REA.A – Marie Skłodowska-Curie Actions & Support to Experts within the project 101182948 — RETROTRAFO — HORIZON-MSCA-2023-SE-01.

The authors also gratefully thanks Dr. Michal Krbal and acknowledge the Centre for Research and Utilization of Renewable Energy, Brno University of Technology, Czech Republic, for providing the CVVOZEPowerLab Research Infrastructure.

This work was also supported by the Erasmus+ mobility programme under the institutional code SK KOSICE03, which enabled the research stay at the Faculty of Electrical Engineering and Communication, Brno University of Technology.

References

[1] Y. Ge *et al.*, "Positive streamer characteristic at the interface of TiO2 nanofluid/pressboard under lightning impulse voltage," *2017 IEEE 19th International Conference on Dielectric Liquids (ICDL)*, Manchester, UK, 2017, pp. 1-4, doi: 10.1109/ICDL.2017.8124623.

[2] Y. Zhou, S. Sui, J. Li, Z. Ouyang, Y. Lv, C. Li and W. Lu, "The effects of shallow traps on the positive streamer electrodynamics in transformer oil based nanofluids," *Journal of Physics D: Applied Physics*, vol. 51, no. 10, p. 105304, 2018, doi: 10.1088/1361-6463/aaab2a.

[3] J. G. Hwang, F. O'Sullivan, M. Zahn, O. Hjortstam, L. A. A. Pettersson and R. Liu, "Modeling of Streamer Propagation in Transformer Oil-Based Nanofluids," *2008 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Quebec, QC, Canada, 2008, pp. 361-366, doi: 10.1109/CEIDP.2008.4772777.

[4] J. G. Hwang, M. Zahn, F. M. O'Sullivan, L. A. A. Pettersson, O. Hjortstam and R. Liu, "Effects of nanoparticle charging on streamer development in transformer oil-based nanofluids," *Journal of Applied Physics*, vol. 107, no. 1, p. 014310, 2010, doi: 10.1063/1.3267474.

[5] Y. Lv, Q. Wang, Y. Zhou, C. Li, Y. Ge and B. Qi, "Effect of Fe_3O_4 nanoparticles on positive streamer propagation in transformer oil," *AIP Advances*, vol. 6, no. 3, p. 035110, 2016, doi: 10.1063/1.4944487.

[6] M. Rafiq, Y. Lv and C. Li, "A review on properties, opportunities, and challenges of transformer oil-based nanofluids," *Journal of Nanomaterials*, vol. 2016, Art. no. 8371560, 23 pages, 2016, doi: 10.1155/2016/8371560.

Morphological and Syntactic Analysis of Synthetic Questions Generated by LLM Models for Slovak Wikipedia

[1]*Ing. Kristián SOPKOVIČ (2nd year),*
Supervisor: [2]*doc. Ing. Daniel HLÁDEK, PhD.*

[1,2]Deptartment of computer Networks, FEI, Technical University of Košice, Slovak Republic

[1]kristian.sopkovic@tuke.sk, [2]daniel.hladek@tuke.sk

***Abstract*—This work focuses on morphological and syntactic analysis of questions generated by large language models for Slovak Wikipedia compared to the reference SKQUAD dataset. We analyze seven different large language models from the Gemini family, examining linguistic characteristics of generated questions at the levels of morphology, syntax, and question typology. We generated approximately 17,500 synthetic questions from 500 Wikipedia contexts. Experimental results show that synthetic questions exhibit significantly higher morphological richness (1.49–1.56 vs 1.37, $p < 0.001$) and syntactic complexity (average tree depth 4.1–4.8 vs 3.88, $p < 0.01$) compared to original questions. Most critically, we identify statistically significant differences in question type distribution ($p < 0.001$), where synthetic questions prefer "WHAT KIND" (35–40%) and "WHICH" (19–28%) questions at the expense of "WHAT" questions (10–14% vs 21.6%).**

***Keywords*—question generation, large language models, morphological analysis, SKQUAD, syntactic complexity, Slovak NLP**

I. Introduction

Question generation using large language models (LLMs) represents a promising research area in natural language processing, particularly for morphologically rich languages such as Slovak. While numerous datasets and studies evaluating the quality of generated questions exist for English, this area remains in its early development stage for Slovak.

Slovak belongs to inflectional languages with rich morphology, including seven cases, three genders, and a complex system of verb forms [4]. This linguistic complexity presents specific challenges for LLM models in generating meaningful questions. The SKQUAD dataset [1] is the first manually annotated dataset of questions and answers in Slovak, consisting of over 91,000 factual questions from various fields. However, creating such datasets is time-consuming and financially demanding. Using LLM models to generate synthetic questions could significantly reduce these costs, but thorough analysis of the linguistic quality of such generated questions is necessary.

Our motivation stems from the need to understand the extent to which synthetic questions generated by LLM models differ from human questions. We address the following research questions: **RQ1:** How does the morphological richness and part-of-speech distribution of synthetic questions compare to human-generated questions? **RQ2:** What are the differences in syntactic complexity between synthetic and human questions? **RQ3:** Are there systematic biases in the distribution of question types in synthetic data?

II. Related Work

Recent studies have shown that LLM models tend to produce less diverse outputs compared to human texts [2]. Guo et al. developed a comprehensive framework for evaluating linguistic diversity in three dimensions: lexical, syntactic, and semantic. Their research showed that models exhibit lower syntactic diversity in creative tasks. The Benchmarking Linguistic Diversity study [2] proposed metrics based on graph kernels for quantifying syntactic diversity using dependency trees. They showed that LLM models achieve near 100% accuracy in generating syntactically plausible sentences but capture only 35–75% of human text syntactic diversity.

In the context of morphologically rich languages, this issue is even more pronounced. Munková et al. [4] demonstrated the use of POS tags and frequency tagsets for evaluating machine translation quality for Slovak. Their methodology, based on association rules and morphological complexity, provides a foundation for our analysis.

Guo et al. [3] examined the decline in linguistic diversity during recursive training of language models on synthetic data, revealing consistent decline in lexical, syntactic, and semantic diversity. Kirk et al. [5] analyzed the impact of RLHF on LLM model generalization and diversity, finding that RLHF reduces output diversity. To our knowledge, this is the first systematic linguistic analysis comparing human and LLM-generated questions for Slovak.

III. Methodology

A. Dataset and Experimental Setup

Our experimental dataset consists of 500 contexts randomly selected from Slovak Wikipedia articles across diverse topics (history, geography, science, culture, sports). For each context, we generated approximately 5 questions using seven different Gemini models (gemini-2.5-pro, gemini-2.0-flash-lite, gemini-3-pro-preview, gemini-3-flash-preview, gemini-2.5-flash-lite, gemini-2.0-flash-001, gemini-2.5-flash), resulting in 2,495–2,500 questions per model (total: approximately 17,500 synthetic questions).

Generation procedure: We used the following prompt for all models: *"Generate 5 diverse questions based on the following Wikipedia paragraph. Questions should be answerable from the text and cover different aspects of the content."* Temperature was set to 0.7, and top-p to 0.9 for all models to ensure sufficient diversity while maintaining coherence. As our reference dataset, we use a subset of SKQUAD [1] corresponding to the first 500 articles, containing 2,460 original questions created by human annotators.

B. Morphological Analysis

For morphological analysis, we use the TreeTagger tool adapted for Slovak, achieving approximately 95% accuracy on Slovak POS tagging. We analyze: (1) **Morphological richness** = unique POS tag combinations / total tokens; (2) **Part-of-speech distribution** using the universal POS tagset scheme (UPOS); (3) **Case distribution** across seven Slovak grammatical cases (nominative, genitive, dative, accusative, locative, instrumental, vocative).

C. Syntactic Analysis

We evaluate syntactic complexity using dependency trees created by the Stanza tool (version 1.4.0) [6] with the Slovak UD model, achieving approximately 90% unlabeled attachment score (UAS). We measure: (1) **Average dependency tree depth** - the maximum distance from root to leaf node, averaged across questions; (2) **Average dependency distance** - the average absolute difference in word positions between dependent and head tokens.

D. Question Typology and Statistical Testing

We classify questions into categories based on their initial interrogative words using automatic pattern matching against Slovak interrogative markers: ČO (WHAT), AKÝ (WHAT KIND), KTORÝ (WHICH), KEDY (WHEN), KTO (WHO), AKO (HOW), KDE (WHERE), KOĽKO (HOW MANY), PREČO (WHY), and others. We use independent samples t-tests for comparing continuous variables and chi-squared tests for categorical comparisons, with significance level $\alpha = 0.05$. Effect sizes are reported using Cohen's d for continuous variables and Cramér's V for categorical comparisons.

IV. Results and Analysis

A. Morphological Richness

Table I presents morphological richness values. All synthetic datasets exhibit significantly higher morphological richness (1.49–1.56) compared to original SKQUAD (1.37). Statistical testing confirms these differences are significant ($t = 8.34$, $p < 0.001$, Cohen's $d = 0.52$), indicating a medium-to-large effect size. This suggests that LLM models generate morphologically more complex questions, utilizing a broader variety of word forms.

B. Part-of-Speech and Case Distribution

POS tag analysis revealed systematic differences (Table II). The table shows the percentage of questions containing at least one instance of each POS tag. Synthetic questions more frequently contain determiners (21–28% of questions vs 11%), prepositions (32–40% vs 22%), and adjectives (40–41% vs 24%), while they less frequently contain proper nouns (16–20% vs 28%). These differences reflect LLM models' preference for more explicit grammatical constructions and more generic descriptive language over specific named entities.

TABLE I
Morphological Richness with Statistical Significance

Model	Richness	p-value
SKQUAD (reference)	1.374	–
gemini-2.5-pro	1.557	< 0.001
gemini-2.0-flash-lite	1.485	< 0.001
gemini-3-pro-preview	1.516	< 0.001
gemini-3-flash-preview	1.543	< 0.001
gemini-2.5-flash-lite	1.534	< 0.001
gemini-2.0-flash-001	1.494	< 0.001
gemini-2.5-flash	1.547	< 0.001

TABLE II
Distribution of Selected POS Tags (Percentage of Questions Containing Tag)

POS	SKQUAD	g-2.5-pro	g-3-flash
DET	11.4	25.5	27.7
NOUN	52.8	88.0	90.1
VERB	19.8	26.4	25.5
ADJ	24.3	40.2	40.6
ADP	21.8	39.7	40.8
PROPN	27.5	19.2	17.9

TABLE III
Case Distribution (Percentage of Questions Containing Case)

Case	SKQUAD	g-2.5-pro	g-3-flash
Nominative	53.6	48.2	45.8
Genitive	31.4	63.7	57.2
Dative	8.7	12.4	11.8
Accusative	47.8	51.9	49.3
Locative	29.6	53.8	48.1
Instrumental	15.2	22.6	19.7
Vocative	0.8	0.3	0.4

Case distribution analysis (Table III) reveals fundamental differences. SKQUAD exhibits nominative dominance (53.6% of questions contain nominative nouns), while synthetic questions show a more balanced distribution with significantly higher proportions of genitive (57–64% vs 31%) and locative (42–54% vs 30%). This reflects LLM models' preference for complex prepositional constructions (which typically govern genitive or locative cases) over simple nominal phrases in nominative case.

C. Syntactic Complexity

Synthetic questions achieve significantly higher syntactic complexity (Table IV). Average tree depth is 5–24% higher (4.1–4.8 vs 3.88, $t = 4.67$, $p < 0.01$, Cohen's $d = 0.38$), indicating deeper nested syntactic structures. Average dependency distance is also significantly higher (2.27–2.65 vs 2.23, $t = 3.12$, $p < 0.05$, Cohen's $d = 0.24$), suggesting more

TABLE IV
SYNTACTIC COMPLEXITY WITH STATISTICAL SIGNIFICANCE

Model	Tree Depth	Dep. Dist.	p-value
SKQUAD	3.878	2.235	–
gemini-2.5-pro	4.808	2.618	< 0.001
gemini-2.0-flash-lite	4.100	2.275	< 0.05
gemini-3-pro-preview	4.565	2.473	< 0.01
gemini-3-flash-preview	4.654	2.543	< 0.01
gemini-2.5-flash-lite	4.521	2.514	< 0.01
gemini-2.0-flash-001	4.356	2.389	< 0.05
gemini-2.5-flash	4.742	2.647	< 0.001

TABLE V
DISTRIBUTION OF QUESTION TYPES (TOP 6, %)

Type	SKQUAD %	g-2.5-pro %	g-3-flash %
WHAT	21.6	10.4	10.3
WHAT KIND	21.4	35.0	32.3
WHICH	12.9	19.0	27.9
WHEN	12.5	3.9	4.2
WHO	9.0	6.2	5.4
WHY	1.0	4.1	3.9
Top 6	78.4	78.6	84.1
Other	21.6	21.4	15.9
χ^2	–	585.76	569.06
p-value	–	< 0.001	< 0.001

complex syntactic relationships between words with greater separation between dependents and their heads.

D. Question Typology

The chi-squared test revealed statistically significant differences in question type distribution for all models compared to SKQUAD. Key observations are as follows: SKQUAD shows a balanced distribution with "WHAT" (21.6%), "WHAT KIND" (21.4%), and "WHICH" (12.9%) as dominant types, while synthetic questions exhibit a significant shift toward "WHAT KIND" (32–35%) and "WHICH" (19–28%) at the expense of "WHAT" and temporal "WHEN" questions. Full counts and statistics are presented in Table V.

V. DISCUSSION

A. Morphological and Syntactic Patterns

Higher morphological richness (Cohen's $d = 0.52$) indicates LLM models utilize a broader spectrum of morphological forms. The increase in genitive and locative cases aligns with findings by Munková et al. [4], reflecting preference for prepositional constructions (e.g., "v meste"/"in the city") over simple nominal phrases. This may stem from training data characteristics, explicit prompting for diversity, or inherent model behavior.

Higher syntactic complexity (Cohen's $d = 0.38$) is consistent with Guo et al. [2], who found models produce syntactically correct but structurally uniform outputs. Manual inspection of 100 samples confirmed grammatical correctness despite increased complexity. Model variation suggests architecture and training procedures influence syntactic complexity.

B. Typological Bias and Implications

The 50% reduction in "WHAT" questions and 60% increase in "WHAT KIND" questions represents our most critical finding. This systematic bias likely stems from prompt interpretation, training data bias toward analytical questions, or model preference for semantic disambiguation. This has significant implications: QA models trained predominantly on "WHAT KIND" questions may underperform on simple fact retrieval tasks common in real applications.

We recommend: (1) targeted prompting for balanced distributions, (2) post-processing filtering, (3) hybrid human-synthetic approaches. **Limitations:** single domain (Wikipedia), no human evaluation, no downstream validation, tool accuracy 90–95%, only Gemini models tested.

VI. CONCLUSION

Analysis of approximately 17,500 synthetic questions from seven large language models reveals: (1) significantly higher morphological richness (1.49–1.56 vs 1.37, $p < 0.001$, $d = 0.52$) with preference for prepositional constructions, (2) higher syntactic complexity (tree depth 4.1–4.8 vs 3.88, $p < 0.01$, $d = 0.38$), and (3) critical systematic bias in question types ($p < 0.001$, $V = 0.31$) with 50% fewer "WHAT" and 60% more "WHAT KIND" questions. While LLMs generate linguistically sophisticated Slovak questions, systematic biases limit direct substitutability of synthetic for human data. We recommend cautious hybrid approaches with targeted prompting to maintain typological diversity. Future work will focus on extending this analysis to additional LLM families and domains, with the goal of developing a domain-adaptive synthetic data augmentation pipeline for Slovak question answering as part of the author's ongoing PhD research.

ACKNOWLEDGMENT

This study was partially funded by the Ministry of Education, Research, Development and Youth of the Slovak Republic under the projects KEGA 049TUKE-4/2024 & VEGA 1/0685/26, by the Slovak Research and Development Agency under the projects APVV-22-0414 & APVV-22-0261 and also from „Synthetic Data Augmentation and LLM-Based Instruction Generation for Robotic Control" project under the joint call of TUKE and Óbuda University, within the "TUKE–OU Young Researchers' Grant" research and development program.

REFERENCES

[1] D. Hládek, J. Staš, J. Juhár, and T. Koctúr, "Slovak dataset for multilingual question answering," *IEEE Access*, vol. 11, pp. 32869–32881, 2023.

[2] Y. Guo, G. Shang, and C. Clavel, "Benchmarking linguistic diversity of large language models," *Transactions of the Association for Computational Linguistics*, vol. 13, pp. 1507–1526, 2025.

[3] Y. Guo, G. Shang, M. Vazirgiannis, and C. Clavel, "The curious decline of linguistic diversity: Training language models on synthetic text," in *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 3589–3604, 2024.

[4] D. Munková, M. Munk, Ľ. Benko, and J. Šťastný, "MT evaluation in the context of language complexity," *Complexity*, vol. 2021, Article ID 2806108, 2021.

[5] R. Kirk, I. Mediratta, C. Nalmpantis, J. Luketina, E. Hambro, E. Grefenstette, and R. Raileanu, "Understanding the effects of RLHF on LLM generalization and diversity," in *The Twelfth International Conference on Learning Representations*, 2024.

[6] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A python natural language processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 101–108, 2020.

Dynamic Routing QoS-Guaranteed Software Defined Network Architecture

[1]*Erika Abigail KATONOVÁ (3rd year),*
Supervisor: [2]*Peter FECIĽAK*

[1,2]Dept. of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]erika.abigail.katonova@tuke.sk, [2]peter.fecilak@tuke.sk

***Abstract*—Traditional routing protocols fail to guarantee QoS in dynamic networks due to reactive management and legacy hardware limitations. This paper proposes a hybrid architecture with ESP32 REST-to-CLI conversion, hybrid topology discovery, ML-enhanced routing (reinforcement learning, A*/Dijkstra heuristics), NetFlow/ELK flow detection, and failure prediction. The system enables dynamic routing ensuring low latency and reliable throughput without hardware replacement.**

***Keywords*—Dynamic routing, machine learning, network architecture, reinforcement learning, QoS**

I. Introduction

The digital world is changing quickly. The growth of the Internet of Things (IoT), powerful cloud platforms, and applications that rely on real-time data has put increasing pressure on traditional network systems. Older routing methods such as OSPF, BGP, and static configurations were built for networks with predictable traffic and stable structures [1]. As a result, they often struggle to maintain consistent Quality of Service (QoS) in today's fast and unpredictable environments [1]. These systems are mainly reactive, meaning they can respond to link failures, but they are not designed to anticipate congestion or automatically adjust routes based on complex, real-time network data.

To overcome these limitations, this paper introduces a new multi-layered network architecture that shifts management from a static and reactive approach to one that is proactive. The proposed system combines machine learning, real-time analytics, and detailed simulation tools to create a network control system that can optimize itself. One key contribution of this research is a hardware-independent interoperability layer. This layer uses a custom-built module that converts JSON-based REST API requests into traditional Command Line Interface (CLI) commands through ESP32 microcontrollers [2]. This allows modern, software-defined control of older network hardware that does not support RESTCONF.

The architecture also improves network reliability through a hybrid monitoring system that blends distributed and centralized methods. By maintaining a real-time map of the network, the system can run advanced routing algorithms that consider more than just bandwidth or hop count. In addition, the use of machine learning allows the system to analyze traffic patterns over time, helping detect potential congestion or hardware issues before they become serious problems.

[1]https://ipcisco.com/lesson/ospf-vs-bgp/

By connecting legacy infrastructure with modern predictive technologies, this research presents a practical solution for upgrading network management. The following sections describe the system's design, the method used for intelligent path selection, and the simulation results that demonstrate its ability to maintain consistent transmission quality in complex and dynamic network environments.

II. Proposed solution

Before describing the proposed architecture, depicted in Fig. 1 it is important to understand the main weaknesses in current network management systems that this research seeks to address. Many organizations still rely on older routers that do not support modern management protocols such as RESTCONF. Replacing this hardware is often expensive and disruptive. At the same time, traditional routing protocols like OSPF and BGP are largely reactive, meaning they respond to problems only after they occur, such as when a link fails, rather than anticipating congestion or performance degradation in advance. In addition, most routing decisions are based on simple metrics like shortest path or available bandwidth. These limited criteria do not account for real-time performance factors such as delay, jitter, or the specific quality requirements of different applications, making it difficult to maintain consistent service levels in dynamic network environments [3].

Another major limitation is the lack of a detailed and accurate representation of the overall network topology. Without a complete and up-to-date view of all nodes, links, and performance parameters, it is difficult to make informed routing decisions. Effective optimization requires working with comprehensive network information, rather than isolated or partial data. A system that maintains a full, real-time model of the network can evaluate multiple variables simultaneously and make more intelligent decisions.

III. Components of architecture

The solution is composed of ten interconnected modules that manage the lifecycle of a network packet, from hardware-level data acquisition to machine-learning-driven path optimization. All of them are depicted in Fig. 1 below on the scheme, showing the communication between them as well as a short description of how the individual modules operate and the functions they provide. The following subsections will introduce the modules one by one and their importance in this whole solution.

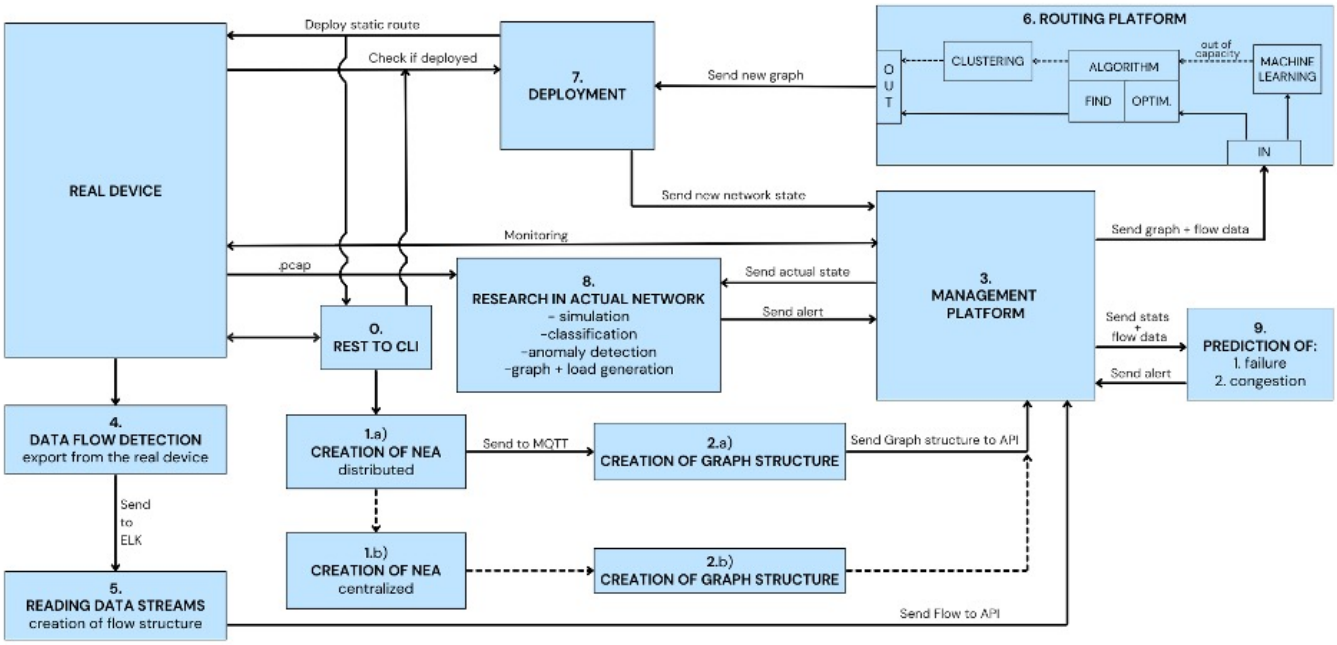


Fig. 1. Proposed architecture

A. CLI converter

The first module marked as 0. is the CLI-to-REST API converter, which purpose is simple - it connects modern network management systems with older legacy routers that do not support modern protocols. The module is built around an ESP32 microcontroller combined with an RS232-to-TTL interface, which allows it to connect directly to a router's console port [2]. This makes it suitable for devices running older operating systems, such as Cisco IOS versions earlier than 16.0.0, where RESTCONF or NETCONF are not available.

The converter accepts requests in JSON format, structured using YANG models, and translates them into the specific CLI commands required by the target device. Instead of relying on in-band communication through the main network, it operates over a direct serial connection. This approach means management access remains available even if the data plane fails. By adding this translation layer, older routers can be managed in the same way as newer, software-defined devices [4]. In effect, the module allows legacy hardware to participate in a centralized, data-driven control system.

B. Node-edge advertisement creation

The NEA Module is the following module after the responsible for real-time topology discovery and the maintenance of the global network graph. To ensure high availability and data integrity, the module utilizes a hybrid approach consisting of two distinct discovery mechanisms:

Accurate, real-time network topology discovery forms the backbone of effective routing and management in dynamic environments, yet legacy systems often fail to provide automated, comprehensive mapping. The NEA module generates structured JSON objects representing nodes (vertices with attributes like ID, status, interfaces) and edges (links with metrics such as bandwidth, latency, MTU). To ensure high availability and data integrity, the module utilizes a hybrid approach consisting of two distinct discovery mechanisms:

1) **Distributed**: In the distributed approach, each router connects to a dedicated ESP32 microcontroller via RS232 serial interface to its console port. The ESP32 operates on configurable intervals, typically every 30-60 seconds, issuing standard diagnostic commands such as *show ip interface brief* to capture node details like operational status and capabilities. Raw console outputs are parsed through regex patterns and predefined JSON templates, giving us two distinct payloads: a node-specific JSON detailing the router's attributes and an edge JSON for interconnections. These structured data packets then publish asynchronously via MQTT protocol to a central broker.
2) **Centralized**: The centralized approach employs a server that begins with SSH access to a router, executes *show cdp neighbors* to identify adjacent devices, and recursively follows the topology via SSH sessions, progressively aggregating node and edge data into a unified JSON graph structure. This complete representation pushes directly to the control platform's endpoint, eliminating the MQTT layer.

The node-edge advertisements are then passed directly into the graph processing component, where they are used to calculate routing paths using machine learning techniques. This approach goes beyond traditional methods that rely only on bandwidth or hop count, allowing routing decisions to consider a wider range of performance factors.

C. Graph structure

The hybrid NEA module provides node and edge data as separate JSON fragments, which must be combined into a single network graph to be used by other parts of the architecture. This assembly process adapts to the discovery method used, balancing efficiency and real-time accuracy. In

[2]https://www.cisco.com/c/en/us/products/collateral/routers/829-industrial-router/white-paper-c11-743674.html?utm_source=chatgpt.com

the distributed mode, a server subscribes to MQTT topics to receive these fragments from ESP32 agents as they arrive. The server collects the data, checks it for consistency (such as matching edges from both ends), and merges it into one complete graph. This graph uses a simple structure with nodes listed by ID and status, edges by connections and metrics, and a timestamp.

The final graph then sends to the control platform through a REST API call to a specifcic endpoint. This allows other modules to work with an up-to-date view of the network. The process runs on a schedule or when changes occur, handling additions or failures smoothly.

The centralized mode skips this step. The SSH/CDP process already builds the full graph during discovery and sends it directly to the edpoint. The module turns raw data into a foundation for smart routing and QoS control. The next subsection covers the control platform that manages these graphs.

D. Management platform

The module marked as 3. in the architecture is a control platform, that acts as the central hub of the architecture, collecting data from discovery modules and coordinating all management functions. It stores and processes key network information, including topology graphs, device configurations, active data flows, monitoring metrics, and baseline profiles. This platform also gathers notifications from detection systems and serves as the main access point for all data through REST API endpoints. For handling unstructured data like JSON graphs and logs, it uses MongoDB as the database backend, which supports flexible schemas and high scalability.

The platform provides real-time visualization of the network topology, allowing administrators to view the current graph structure and track changes. When a new data flow registers, it updates its records and forwards the latest graph along with flow details to the routing platform for path computation. After the routing platform responds with an optimized graph (now including the new path), the control platform receives and stores this updated version, ensuring all modules operate on synchronized state. This design ensures reliable data management and quick adaptation to network events.

E. Data flow detection

Module marked as 4. in the architecture, is the Data flow detection module, which identifies active network traffic patterns to enable QoS-aware routing decisions. Edge routers in the topology run NetFlow collectors, which capture flows based on key identifiers: source and destination IP addresses, protocol type, and port numbers. These collectors also gather transmission statistics, such as packet counts, byte volumes, and duration, providing a clear picture of ongoing communications.

The collected NetFlow records export to a Logstash server, which parses the raw data and stores it as structured JSON objects in an Elasticsearch database. Administrators can then visualize these flows through Kibana dashboards, revealing traffic volume and application patterns across the network.

A dedicated server application continuously monitors the Elasticsearch database for new or changed flows (module 5. in the architecture). Upon detecting unregistered traffic, the application extracts essential details and sends them via REST API to the control platform's endpoint. This triggers flow registration, updates the topology visualization, and forwards the information to the routing platform for optimal path computation.

This module ensures the architecture maintains complete visibility into live traffic without intrusive deep packet inspection. By automating flow discovery and integrating seamlessly with the control layer, it supports real-time adaptation to changing network demands while preserving performance. The next subsection examines how the routing platform processes these flows against the current topology.

F. Routing platform

The routing platform, marked as 6. in the architecture, takes the current network graph and new flow data from the control platform to compute QoS-guaranteed routes, implementing dynamic routing without OSPF/BGP convergence delays. It features three modules: machine learning, algorithms, and clustering. The machine learning module uses reinforcement learning for self-improving capacity optimization and custom heuristics (latency, loss, reliability) to enhance A*/Dijkstra pathfinding. The algorithmic module finds flow-aware paths and resolves congestion via the Optim submodule. The clustering module groups large topologies into OSPF-like areas, computing paths between clusters for scalability. It outputs an updated graph with the new path back to the control platform for deployment, ensuring QoS through integrated ML and algorithms. The next subsection covers configuration deployment.

G. Configuration deployment

The configuration deployment module, number 7. in architecture, receives the updated graph structure from the routing platform, which now includes the newly computed paths for active data flows. It verifies the current network state against the proposed changes by querying router configurations through the REST-to-CLI converter, ensuring no conflicts or invalid routes exist before proceeding. The module then translates the optimal paths into specific CLI commands and deploys them sequentially across affected routers via console access.

Once deployment completes, the module captures confirmation outputs from each device, validating successful installation through parsing of *show ip route* results. A success or failure status, along with execution logs, returns via REST API to the control platform, which updates its topology records and notifications accordingly. This dynamic routing approach installs persistent routes that adapt to flows without full protocol reconvergence, maintaining uninterrupted QoS during updates.

The process operates with rollback capability: if verification fails or partial deployments occur, it reverts changes automatically, what ensures network stability. The next subsection explores research components and extra parts of the whole architecture for continuous network improvement.

H. Research and network improvement

The research components, in the module 8. and prediction module 9., enable continuous self-improvement of the network through simulations, data analysis, and predictive modeling,

addressing long-term evolution beyond immediate operations. These tools generate insights that refine machine learning models, detect anomalies, and prevent failures, ensuring sustained QoS as traffic patterns evolve.

Simulations form the core of experimentation. The ns-3 simulator recreates the real topology and data flows at a 1:1 scale, allowing time-forward execution to predict future issues like jitter, delay, or packet loss based on packet statistics [5]. Mininet complements this by generating realistic traffic captures saved as .pcap files, which serve as training datasets for machine learning models and neural networks [6]. A standard profile emerges from captured real data or Mininet outputs, against which new flows compare via unsupervised outlier or clustering methods. This detects anomalies like attacks, faulty devices, or degrading media quality, flagging issues for manual review. For large-scale testing, tools generate complex graph structures mimicking real topologies and synthetic data flows with realistic metadata (packet size, frequency, addressing), supporting simulations without relying on live production traffic.

Finally, the failure and congestion prediction module, module 9. in architecture, employs recurrent neural networks (RNNs) on time-series datasets from the control platform. It identifies temporal patterns signaling impending overloads or degrading components, sending notifications to trigger path recomputation, traffic diversion, or administrator alerts, preventing disruptions proactively.

Together, these research elements close the feedback loop, transforming the architecture into robust platform capable of autonomous state discovery and proactive path optimization.

IV. Development Progress Overview

Research has reached different stages of maturity across its ten-module architecture. Five core components have been fully implemented and tested in a controlled environment, forming a functional prototype. The last three modules remain as a work in progress validated through simulations and requirements analysis.

CLI Converter: This module is fully implemented as ESP32 prototype on Olimex PoE board. It successfully replaces most RESTCONF functionality via CLI for any Cisco router with console access (IOS 15.x validated), using ESP-IDF C framework with UART interface. Performance metrics were gathered from Postman, and show the following: GET empty routes 6.6s, PUT creating 2 static routes 14.61s, existing routes with 2 paths same as empty (proving CPU not bottlenecked, just CLI timing gaps of 500ms post-'enter'). Advantages are low-cost, rack-compact, extensible macros, ideal for testing networks avoiding RESTCONF hardware costs where user DB/password config unnecessary. Proven proof-of-concept demonstrating RESTCONF capability extension to legacy routers.

The components for creation of **Node-Edge Advertisement** and **Graph Structure** were implemented as a client-server prototype, which implements both functions as integrated workflow. ESP32 serial clients and RESTCONF/YANG clients collect router data and serialize it as node/edge JSON fragments, which the server then aggregates into complete topology graphs exposed through REST API. Tested across virtual topologies and hybrid real/virtual environments, this system reliably detects devices, extracts configurations, and produces consistent graph models while resolving challenges like ESP32 memory constraints and DB9 connector issues.

The **Data flow detection module** is fully implemented using Elastic Stack (Elasticsearch, Logstash, Kibana), Python scripting, and MQTT broker integration, deployed in VirtualBox with Cisco CSR1000v generating flows and Ubuntu server processing via UDP 9995. Evaluation confirms real-time detection of new flows (ICMP ping, TFTP port 69, HTTP port 80), automatic removal of inactive flows after 30s inactivity based on @timestamp, and MQTT notifications for flow start or end events. Logstash transforms data, Kibana visualizes traffic, Python script polls Elasticsearch every 10s. The evaluation confirms a reliable bidirectional flow tracking, with no visible detection latency. The module is also dockerized for easy deployment.

The **Research module** integrates validated simulation and prediction tools for network analysis.

First off the **Simulation** part of the module is fully implemented as a desktop application, which generates structured JSON topology files and visualizes them as node/edge graphs for scenario validation. The application generates realistic data flows between separate network instances simulating VoD, live video/audio streaming, and VoIP calls with near-identical characteristics to real services (packet sizes, timing, volume differences minimal). The evaluation confirms generated topologies match predefined structures and flows replicate real service behavior. This simulation platform also simplifies experimental network scenario creation and research validation.

The next defined was **Classification** part, which is implemented as a bi-directional LSTM architecture, which classifies 8 service types (HTTP, HTTPS, VPN, VoIP, etc.) across 3 subnets (internal, demilitarized and external). The best-performing model, initialized with seed 42, achieved a test accuracy of approximately 87.2% and a weighted F1-score of 0.869; notably, it successfully avoided the significant overfitting observed in other models, which saw performance drops to 50–60% on unseen data. While services like HTTP and VPN were classified with over 90% accuracy, SMTP and IMAP proved more challenging, falling below 80%. Ultimately, the model demonstrated a strong capacity to maintain network traffic structures and temporal dynamics.

The **Anomaly** and **Congestion Detection** part is also fully implemented and combines CNN-LSTM autoencoder (traffic anomalies via reconstruction error, complete with database/REST API/web interface) and ARIMA/SARIMA prediction models. The congestion evaluation shows that ARIMA excels in short-term CPU predictions and interface load prediction, SARIMA on CESNET real data shows NRMSE 0.9684, R^2 0.5324 for long-term n_bytes with seasonal trend capture (LSTM/classification failed due to data variability). The evaluation therefore shows the early warning capabilities vs reactive monitoring.

The anomaly detection was also implemented as an additive extra modul, which detects network traffic anomalies by training on historical standard profiles and identifying deviations through reconstruction error. Complete implementation includes flow database for traffic/results storage, REST API for module communication, and web interface for anomaly visualization/management. Overall accuracy was 79-91%, and F1-score 82-91% across experiments, the anomaly class precision was 82%, and recall 98%; normal traffic precision

97%, and recall 79%. High anomaly sensitivity prioritized over precision to ensure security (better false positives than missed threats). Confusion matrix confirmed reliable identification with minimal missed anomalies.

The **Management Platform** is still work in progress, it centralizes all incoming data from discovery modules (NEA/Graph) and provides real-time topology visualization. MongoDB integration and core REST API endpoints (/API/Graph, /API/Flow) implemented. Basic topology storage functional; visualization dashboard and full data orchestration between modules under active development.

The **Routing Platform** and **Deployment module** remain also under active development. These modules represent the final phase to achieve end-to-end "dynamic static routing" with guaranteed QoS, building on validated discovery and prediction components. Current prototype focuses on proven foundation; routing automation completes the architecture.

V. Conclusion

This research proposes a comprehensive network architecture that successfully addresses the limitations of traditional routing protocols like OSPF, BGP, and static configurations, which struggle with dynamic traffic, reactive management, and QoS guarantees in modern environments driven by IoT, cloud computing, and real-time applications. Through ten integrated modules, from the ESP32-based REST-to-CLI converter and hybrid NEA discovery, to ML-enhanced routing, flow processing, and predictions, the system shifts networks from static reactivity to proactive, data-driven intelligence, ensuring consistent transmission quality parameters such as low latency, minimal jitter, and reliable throughput. By enabling cost-effective modernization without full hardware replacement, this work enriches the networking industry with a practical blueprint for hybrid environments.

References

[1] G. A. Amiri and S. Shahidi, "Navigating routing challenges in internet of things (iot): A focus on routing protocols and wireless sensor networks," *Cognizance Journal of Multidisciplinary Studies*, vol. 3, no. 8, pp. 297–304, Aug. 2023.

[2] M. Babiuch, P. Foltýnek, and P. Smutný, "Using the esp32 microcontroller for data processing," in *2019 20th International Carpathian Control Conference (ICCC)*, 2019, pp. 1–6.

[3] R. Garg, R. Budhiraja, A. Gupta, H. Sharma, and S. Thangalazhi, "Real-time network topology optimization using dynamic machine learning adaptation," *International Journal for Research in Applied Science and Engineering Technology*, vol. 13, no. 1, pp. 2010–2018, Jan. 2025.

[4] M. Hussain, N. Shah, R. Amin, S. S. Alshamrani, A. Alotaibi, and S. M. Raza, "Software-defined networking: Categories, analysis, and future directions," *Sensors*, vol. 22, no. 15, p. 5551, Jul. 2022.

[5] L. Campanile, M. Gribaudo, M. Iacono, F. Marulli, and M. Mastroianni, "Computer network simulation with ns-3: A systematic literature review," *Electronics*, vol. 9, no. 2, p. 272, Feb. 2020.

[6] N. Gupta, M. S. Maashi, S. Tanwar, S. Badotra, M. Aljebreen, and S. Bharany, "A comparative study of software defined networking controllers using mininet," *Electronics*, vol. 11, no. 17, p. 2715, Aug. 2022.

The Impact of Photovoltaic Systems on Power System Management

[1]*Kamil ŠEVC (2nd year)*
Supervisor: [2]Marek PAVLÍK

[1,2] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]kamil.sevc@tuke.sk, [2]marek.pavlik@tuke.sk

***Abstract* - The development of the energy sector represents one of the most critical challenges of the 21st century. The global energy industry is currently undergoing a major transformation, primarily driven by the necessity to mitigate climate change and to meet internationally agreed climate and energy targets. The integration of large capacities of renewable energy sources, particularly photovoltaic and wind power systems, into power grids presents a fundamental challenge to system stability. Unlike conventional thermal power plants, which provide grid stability through high-inertia rotating generators, renewable energy sources generate electricity via power electronic converters without mechanical inertia. As a result, rapid changes in power flows, as well as voltage and frequency fluctuations, may occur, potentially threatening the stability of the entire distribution system.**

Keywords - renewable energy sources, photovoltaic systems, distribution networks, electricity generation management.

I. INTRODUCTION

The potential of renewable energy sources is considerable, as they are theoretically capable of covering multiple times the current global energy demand. Renewable sources such as biomass, wind, solar, hydropower, and geothermal energy can provide sustainable energy services based on regularly available and locally accessible resources. The transition towards energy systems based on renewable sources is becoming increasingly feasible due to the continuous decline in their costs, while prices of fossil fuels, particularly oil and natural gas, exhibit significant volatility. Over the past three decades, solar and wind power technologies have experienced rapid market growth, accompanied by decreasing investment costs and lower electricity generation costs, while their technical performance and efficiency have continuously improved.

The costs of fossil fuels and renewable energy sources, as well as their societal and environmental impacts, are evolving in opposite directions. At the same time, economic and policy mechanisms necessary for the large-scale deployment and long-term operation of renewable energy technology markets are being dynamically established. It is becoming increasingly evident that future growth in the energy sector will primarily rely on new technologies utilizing renewable energy sources, complemented to some extent by systems based on natural gas, rather than on traditional oil and coal resources. These trends create new market opportunities for innovation and enable the exploitation of emerging markets supporting renewable energy technologies, further strengthened by increasing public acceptance and governmental support initiatives. The development and deployment of renewable energy sources can contribute to greater diversification of energy markets, ensure long-term sustainable energy supply, reduce both local and global emissions, and simultaneously provide economically attractive solutions for specific energy demands. This is particularly relevant for developing countries and rural areas, where renewable energy deployment can also create new employment opportunities.

II. SOLAR ENERGY

Solar energy is currently among the most cost-effective methods of electricity generation compared to other energy sources, while also being widely available and suitable for a broad range of applications. Although the initial investment costs remain relatively high, the operational and maintenance costs of solar energy systems are low; however, their primary limitation is dependence on weather conditions, which necessitates the integration of energy storage systems and consequently increases overall technology costs. Between 1992 and 2020, the solar energy sector experienced exponential growth, evolving from small-scale applications into a well-established source of electricity generation. Since the development of solar cells in the 1950s, multiple countries have adopted solar energy technologies—initially led by the United States, followed by Japan and Germany, while China currently represents the largest producer of solar-generated electricity [1].

A. Situation in the Slovak republic

In the context of Slovakia, solar energy represents the most widespread source of green energy utilized within the private sector, primarily in the form of solar panels and solar thermal collectors. Until 2010, the development of solar energy in the country progressed slowly, with photovoltaic installations playing only a marginal role in broader national energy strategies. Existing barriers stemmed mainly from technological, financial, and regulatory limitations. However, the introduction of state subsidy schemes in 2010 marked a significant turning point, resulting in a substantial increase in both electricity production and consumption from solar energy, as well as a rapid growth in installed photovoltaic capacity. Since 2010, electricity production and consumption from solar sources, together with installed PV capacity, have shown

continuous growth. Although biomass possesses the highest technical potential due to favorable natural conditions, solar energy represents the overall largest renewable energy potential in Slovakia [2].

B. Photovoltaic power potential

It is also important to highlight a critical fact: when considering the installation of a photovoltaic power plant, for example in a household, geographical conditions within Slovakia play a key role. As illustrated in Fig. 1, it is possible to clearly identify which locations in the country have high or low potential for electricity generation from photovoltaic panels, depending on whether the local solar radiation intensity is high or low.

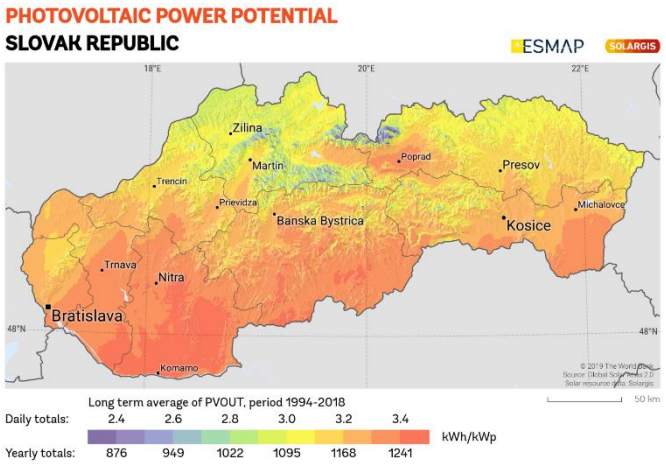


Fig. 1. Map of photovoltaic potential in the Slovak Republic [3]

III. GRID CONNECTION REQUIREMENTS

The issues related to the management, operation, and development of the distribution system are governed by internal documents and technical conditions issued by the system operator, depending on the specific region in which individual distribution companies operate.

In contemporary electricity generation and control, the structure of the power system is evolving towards a safer, more efficient, and technologically advanced smart grid concept compared to the traditional power system. To achieve this objective, smart grids are primarily designed at the level of integrated installations incorporating renewable energy sources. Emphasis is placed on enhancing the reliability of the power system by considering the impacts of connecting wind and solar sources to the grid through the application of advanced intelligent control strategies. Various modelling approaches are employed to achieve the recommended grid control methodology. The implementation of these approaches, together with the definition of general conditions and key parameters, leads to an overall improvement in power system reliability when integrating renewable energy sources into the distribution network [4].

At the substation level, MV/LV transformers equipped with voltage regulation functionality are used to control voltage within the distribution area between the medium-voltage (MV) and low-voltage (LV) networks. In the past, most distribution system operators utilized voltage regulators primarily as a control mechanism to compensate for voltage fluctuations in the MV network and to stabilize the voltage in the LV network at an approximately constant level. With the increasing integration of renewable energy sources into the power system, distribution system operators in some countries have begun implementing power electronics and software-based applications to adjust voltage levels in the LV network depending on the degree of renewable energy penetration and the current power flow at the substation. This approach is intended to suppress voltage rise in the LV network in situations where electricity is injected from renewable energy sources [5].

A. Qualitative Conditions for Connection in Eastern Slovakia

One of the key parameters is the voltage rise limit caused by the operation of connected generation units. Under the most unfavorable operating conditions, the voltage increase at the point of connection must not exceed 3% for sources connected to low-voltage (LV) distribution networks, compared to the condition without connected sources. It is necessary to mention several precise definitions specified in regulation [6]. The document was cited due to the regional focus of this study on Eastern Slovakia, where Východoslovenská distribučná, a.s. acts as the distribution system operator.

The voltage rise resulting from the operation of connected sources must not exceed 3% at the point of connection, in the worst-case scenario, for sources connected to LV systems, compared to the voltage level when no sources are connected. This condition is expressed by the following equation:

$$\Delta u_{nn} b \leq 3\% \tag{1}$$

If there is only one point of connection in the LV network, condition (1) may be applied to determine the maximum connectable apparent power of the source S_{Amax} as:

$$S_{Amax} \leq \frac{3\% \cdot S_{kV}}{\cos(\psi_{kV} - \varphi)} \leq \frac{S_{kV}}{33 \cdot \cos(\psi_{kV} - \varphi)} \tag{2}$$

where the short-circuit power is defined as $S_{kV} = 0.9 S''_{kVmax}$, and ψ_{kV} is the short-circuit impedance angle of the system at the point of connection. If there is only one connection point in the network, the 3% voltage rise criterion will always be satisfied provided that the short-circuit power ratio k_{k1} for sources connected to the LV system meets the following condition:

$$k_{k1nn} \geq 33 \tag{3}$$

In practice, for a given maximum connectable apparent power S_{Amax}, the voltage rise at the point of connection is determined according to:

$$\Delta u_{AV} = \frac{S_{Amax} \cos(\psi_{kV} - \varphi)}{S_{kV}} \tag{4}$$

The voltage change in the system during source switching can be estimated based on the ratio between the short-circuit power of the distribution system operator's network S_{kV} and the rated apparent power of the source S_{nE}, according to:

$$\Delta u_{max} = k_{imax} \frac{S_{nE}}{S_{kV}} \tag{5}$$

where the coefficient k_{imax}, referred to as the "maximum switching inrush factor," represents the ratio between the maximum current occurring during the switching transient I_a and the rated current of the generator or equipment:

$$k_{imax} = \frac{I_a}{I_{nG}} \quad (6)$$

To minimize adverse impacts of generation sources on the distribution system, simultaneous switching of multiple generators at a single connection point should be avoided. A practical solution is the time-staggered switching of individual units. The required interval depends on the induced voltage change: at maximum permissible generator output, a minimum delay of 1.5 minutes is required, while for apparent power up to half of the permissible value, a 12-second interval is sufficient. In cases of very low switching frequency (e.g., once per day), the distribution system operator may allow larger voltage variations during switching, provided that network conditions permit it. Inverters may be switched on only when their terminal side is de-energized.

These requirements form part of the broader technical conditions governing source integration into the distribution network. Their correct application and interpretation are illustrated in the 2023 study conducted by the Department of Electric Power Engineering at the Technical University of Košice [7].

It is also important to mention additional qualitative parameters related to the connection of new generation sources to the distribution network. In accordance with [8], [9], all interconnected components within the distribution network must be electromagnetically compatible. Electrical equipment connected to the distribution network must not generate electromagnetic interference that could impair the operation of other electrical devices and must also exhibit sufficient immunity to disturbances that can be expected within the distribution network. Only such equipment may be connected to the distribution network whose back effects do not negatively affect voltage quality in the network or its users.

IV. Conclusion

The integration of renewable energy sources into modern power systems represents one of the most significant technical and regulatory challenges of the 21st century. Although renewable energy sources offer long-term opportunities for emission reduction, energy security, and energy independence, large-scale integration at high penetration levels remains technically demanding and economically variable without addressing critical technical drawbacks such as voltage instability, limited utilization of energy storage systems, insufficient infrastructure, and disparities in regulatory frameworks. Investments in infrastructure, energy storage innovation, advanced control systems, and harmonized policy frameworks at the global level are essential for achieving a successful energy transition in line with climate objectives. Without such an integrated and comprehensive approach, it will not be possible to ensure the safe, efficient, and reliable integration of renewable energy sources into modern power systems.

It can be stated that aggregation—both technical and commercial—is a key element for the practical utilization of the flexibility created in this way. By interconnecting multiple households equipped with photovoltaic systems, battery storage, and electric vehicles into a coordinated structure, it is possible to achieve a scalable and controllable source of flexibility. Such an arrangement can significantly influence distribution system operation while simultaneously enabling participation in market mechanisms. In this context, the concept of virtual power plants becomes highly relevant. These represent a portfolio of distributed energy resources and flexible loads managed through a centralized platform so that they operate externally as a single aggregated generation or flexibility resource.

In conclusion, it is necessary to highlight the specific characteristics that arise when integrating photovoltaic systems into the power system, as well as innovative solutions aimed at ensuring its stabilization and safe operation. The analysis confirms that the integration of renewable energy sources into power systems represents one of the most critical technical and regulatory challenges of today. While conventional power plants provide system stability through rotating generators with high inertia, photovoltaic sources generate electricity without mechanical inertia, which may lead to voltage and frequency instability. Empirical studies confirm that this structural difference implies that mere expansion of renewable energy sources, without comprehensive supporting measures, does not ensure the long-term sustainability of the energy system. An analysis of technical connection conditions in accordance with European standards and regional distribution company requirements indicates the necessity of precisely defining qualitative parameters for electricity generation sources.

Acknowledgment

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under the contract no. VEGA 1/0627/24.

References

[1] M. K. H. Rabaia, M. A. Abdelkareem, E. T. Sayed, K. Elsaid, K.-J. Chae, T. Wilberforce, and A. G. Olabi, "Environmental impacts of solar energy systems: A review," *Science of The Total Environment*, vol. 754, p. 141989, 2021, DOI: 10.1016/j.scitotenv.2020.141989.

[2] A. Figulová and I. Gondášová, "Energy Transition in Slovakia – Destiny in/and Change," *European Journal of Transformation Studies*, vol. 11, no. 1, pp. 169–187, 2023.

[3] Suri, Marcel *et al.* Global Photovoltaic Power Potential by Country (English). *Energy Sector Management Assistance Program (ESMAP)* Washington, D.C. World Bank Group. [Online]. Available: http://documents.worldbank.org/curated/en/466331592817725242 . [Accessed: 02.02.2026]

[4] M. Colak and S. Balci, "Intelligent techniques to connect renewable energy sources to the grid: A review," in *2021 9th International Conference on Smart Grid (icSmartGrid)*, 2021, pp. 179–185.

[5] C. D. Iweh *et al.*, "Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits," *Energies*, vol. 14, no. 17, p. 5375, 2021.

[6] Východoslovenská distribučná, a.s., *Technické podmienky prevádzkovateľa distribučnej sústavy – Východoslovenská distribučná, a.s.*, effective Sept. 1, 2024. [Online]. Available: https://www.vsds.sk/edso/domov/technicke-info/technicke-podmienky [Accessed: 07.02.2026]

[7] Ľ. Beňa *et al.*, *Analýza pomerov pripojiteľnosti 1F a 3F zdrojov v NN sústave*. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, Katedra elektroenergetiky, 2023.

[8] Directive 2014/53/EU of the European Parliament and of the Council on the harmonisation of the laws of the Member States relating to making available on the market of radio equipment, 2014.

[9] Government Regulation of the Slovak Republic No. 127/2016 on electromagnetic compatibility, 2016.

Analysis of Monitoring Internal Forces in Single-Lap Adhesive Joints Using Magnetic Microwire Sensors

[1]*Miroslav KOMORNÍK (1st year)*
Supervisor: [2]Rastislav VARGA

[1,2] Dept. of Physics, FEI, Technical University of Košice, Slovak Republic
[2]RVmagnetics, a.s., Štúrová 27/A, 040 01 Košice, Slovak Republic

[1]miroslav.komornik@tuke.sk, [2]varga@rvmagnetics.com

***Abstract*—Ensuring the structural integrity of adhesive joints is a key challenge in modern engineering. Since traditional surface-mounted sensors often fail to detect internal flaws, this paper explores the use of glass-coated amorphous magnetic microwires as contactless, non-destructive sensors for single-lap joints. By utilizing the magnetoelastic effect and magnetic bistability, these embedded sensors can provide real-time data on internal force distribution and material fatigue. Importantly, their integration aims to effectively monitor the joint without compromising the mechanical strength of the bond.**

***Keywords*— Adhesive, Force distribution, Joint, Microwire, Health monitoring**

I. Introduction

The structural integrity of adhesively bonded joints is essential in modern aerospace, automotive, and mechanical engineering [1]. While structural adhesives offer significant advantages, such as uniform stress distribution and lightweight assembly, they lack a reliable method for non-destructively detecting hidden microcracks caused by cyclic loading [2], [3].

The primary motivation for addressing this research task is the urgent need for reliable, continuous Structural Health Monitoring (SHM) systems to prevent sudden and catastrophic structural failures [4]. Traditional surface-applied resistance strain gauges are bulky and require physical galvanic connections, making them incapable of penetrating closed adhesive interfaces without compromising the joint [5]. To overcome this barrier, this paper explores the integration of glass-coated amorphous magnetic microwires as innovative contactless sensors [6]. By exploiting nonlinear magnetization dynamics, these sub-millimeter sensors can monitor internal mechanical stresses wirelessly and in real time [7]. This research synthesizes the mechanical, chemical, and electromagnetic principles of this technology to evaluate its practical viability for SHM in single-lap bonded joints [6], [8], [9].

II. Theoretical Mechanics and Stress Analysis of Single-Lap Joint

To understand exactly what the magnetic microwire measures inside the joint, it is necessary to thoroughly deconstruct the mechanics of force transfer in bonded structures [10]. The single-lap joint is structurally the least demanding in terms of manufacturing, yet paradoxically, it faces the greatest mechanical challenges in force transfer [3], [11].

In an ideal analytical model of shear stress under tensile load, it is often assumed that the shear stress is evenly distributed over the entire contact area. However, the reality of continuum mechanics is significantly more complex due to the eccentricity of the applied force [12]. The forces acting on the respective opposing adherends do not lie on the same geometric axis. This shift in the centers of gravity inevitably causes a bending moment to arise in the joint alongside the primary shear stress [10], [12].

The basic equation for this bending moment M in a lap joint can be expressed as the product of the acting force S and the moment arm a, thus $M=S*a$ [13]. This parasitic bending moment forces the bonded materials to deform, rotate, and bend at the overlap region. It generates not only uneven shear stress but also induces so-called peel stresses. Peel stress acts perpendicularly to the plane of the adhesive layer and functions as a normal tensile force that attempts to tear the adherends apart [13], [14]. If this stress concentration is neglected, cohesive or adhesive cracks instantly initiate at the edges of the adhesive [15].

Furthermore, mechanical systems in real operating conditions are rarely exposed exclusively to static loading. For example, in standardized fatigue testing of carbon fiber reinforced polymers (CFRP), cyclic tensile loading is utilized [16]. Under such repeated cyclic stress, the macromolecular network of the adhesive undergoes fatigue processes, necessitating the integration of continuous monitoring systems [17].

A. Evolution of Analytical Models for Adhesive Joints

The foundation of stress distribution analysis in single-lap joints rests on classical analytical models that have evolved significantly over the past century [3], [8]. Early approaches focused on simplified linear-elastic assumptions. Volkersen's pioneering model from 1938 is universally recognized as the first significant mathematical formulation for lap joints, operating under the assumption that the adherends are subjected only to normal stresses, while the adhesive layer

responds exclusively to shear stress. Although mathematically elegant, Volkersen's one-dimensional approach severely underestimated the maximum loads at the edges of the overlap by ignoring the bending effects caused by load eccentricity [8], [12].

This limitation was critically addressed by Goland and Reissner in 1944, who revolutionized the field by introducing bending moments and transverse (peel) stresses into the mathematical equations [10]. Their model mathematically proved that parasitic bending causes severe peel stress concentrations at the very edges of the bonded overlap, which are primarily responsible for initiating cohesive and adhesive failures [13], [15]. To overcome the restrictions of linear elasticity, Hart-Smith (1973) subsequently improved upon the Goland and Reissner paradigm by incorporating elasto-plastic deformation behaviors of the adhesive, which provided a more realistic prediction of ultimate joint strength [3], [8].

In later decades, more complex general sandwich-type models were introduced by Bigwood and Crocombe, and through-thickness shear variations were proposed by Ojalvo and Eidinoff. Today, while Finite Element Analysis (FEA) has largely replaced these closed-form analytical solutions for complex geometries, the classical Volkersen and Goland-Reissner equations remain the gold standard for validating numerical models and understanding the fundamental mechanical gradients that embedded sensors must detect [10], [12].

B. Conventional Structural Health Monitoring Techniques and Their Limitations

I The necessity of recording the structural integrity of adhesively bonded joints to ensure safe load transmission has driven the development of various embedded SHM technologies. The global literature highlights three dominant approaches: piezoelectric sensors, conductive nanocomposites, and fiber optic sensors[5], [19].

Piezoelectric and Acoustic Sensors: Piezoelectric lead zirconate titanate (PZT) sensors are frequently utilized to generate and detect guided ultrasonic Lamb waves or to perform Electro-Mechanical Impedance (EMI) spectroscopy within structures [4]. While PZT sensors excel at capturing dynamic high-frequency acoustic emissions related to crack propagation, their physical integration into thin adhesive layers presents severe mechanical drawbacks. Research indicates that embedding PZT diagnostic films across a composite structure can reduce the elastic modulus of the host material by up to 20% in tension and over 10% in compression, fundamentally weakening the structure they are meant to protect [5].

Conductive Nanocomposites: A rapidly growing area of research involves self-sensing nanocomposites, created by dispersing conductive nanomaterials (such as carbon nanotubes or graphene) directly into the polymer matrix [19]. This allows for the tracking of electrical resistance changes corresponding to fatigue debonding. However, significant challenges remain regarding the uniform dispersion of these nanofillers. In thicker sections, conductivity variations exceeding 22% can occur, leading to poor signal-to-noise ratios and unreliable spatial resolution for strain mapping [19].

Fiber Optic Sensors: Fiber Bragg Gratings (FBG) and Polymer Optical Fibers (POF) represent the most mature and widely documented embedded sensing technologies [20]. They offer excellent immunity to electromagnetic interference and high multiplexing capabilities. Nonetheless, their physical size is a critical flaw when applied to high-performance structural adhesives. Standard optical glass fibers, complete with their protective cladding and coatings, have a total diameter of approximately 145 µm. Literature unequivocally demonstrates that integrating a 145 µm fiber into a thin structural film adhesive acts as a physical barrier [20]. It hinders homogeneous resin flow, forces an artificial thickening of the bond line, and drastically increases the number of trapped air inclusions and pores. Consequently, this physical intrusion significantly reduces the quasi-static tensile shear strength of the joint, violating the primary rule of NDT: the sensor must not initiate the failure [5].

C. Emergence of Magnetic Microwires as a Disruptive SHM Technology

To circumvent the geometrical and mechanical limitations of FBGs and PZTs, recent global research has pivoted toward amorphous, glass-coated magnetic microwires [6]. Fabricated via the modified Taylor-Ulitovsky method, these sensors consist of a ferromagnetic metallic core enveloped in a continuous, biocompatible, and chemically resistant glass coating [22]. FEM optimization studies have confirmed that a microwire with a total outer diameter of just 35 µm and a 15 µm metallic core is the most mechanically durable configuration [9], [21]. This ultra-thin profile allows the adhesive to flow seamlessly around the sensor during curing, utterly eliminating the void-formation issues associated with thick optical fibers [9].

The operational principle of these sensors relies on the magnetoelastic effect and magnetic bistability. Due to the rapid quenching process during manufacturing and the mismatch in thermal expansion coefficients between the metal core and the glass shell, massive internal residual stresses are locked into the microwire [22]. These internal stresses, coupled with alloys exhibiting high positive magnetostriction (e.g., $Fe_{77.5}Si_{7.5}B_{15}$), result in a unique domain structure featuring a single, massive axial domain wall. When subjected to an alternating magnetic field, the microwire undergoes a rapid, single Barkhausen jump—a complete reversal of magnetization [6], [23]. The critical external magnetic field intensity required to trigger this jump is known as the switching field H_{SW} [23].

Because the switching field is highly sensitive to the total mechanical stress tensor applied to the wire, changes in the structural loading of the single-lap joint directly modulate the magnetic response [7], [23]. Researchers have successfully harnessed this phenomenon through the "Enhanced Induction Method". This technique allows for completely contactless diagnostics; an external excitation coil generates a magnetic field that permeates the joint, while a sensing coil detects the sharp induced voltage peaks generated by the domain wall propagation [7].

D. Magnetoelastic Effect and Contactless Sensing

The basis for using magnetic microwires as strain gauges is the magnetoelastic effect. External mechanical loading causes

a change in the magnetic permeability and magnetic anisotropy of the material [22]. A change in the mechanical stress of the microwire (transferred through the single-lap joint) directly reflects in the change of the switching field H_{SW} [7].

The switching field value H_{SW} is directly determined by the total mechanical stress tensor σ. The total mechanical tensile stress is given by the mechanical tensile stress resulting from the manufacturing process σ, stress caused by the different thermal expansion coefficients σ_Δ, and the external mechanical tensile stress σ, rendering $\sigma = \sigma_i + \sigma_\Delta + \sigma_\alpha$.

Measurement is carried out exclusively via an enhanced induction method, enabling completely contactless and wireless diagnostics [9]. An excitation coil generates an alternating magnetic field. Once the intensity reaches the dynamically changing value of H_{SW}, the bistable microwire instantly reverses its magnetization, generating a sharp induced voltage peak in the sensing winding according to Faraday's law [7]. Evaluation electronics monitor the phase and time shift of these sharp voltage peaks. As mechanical stress in the joint increases, the mean value of the microwire's response changes, allowing highly sensitive remote monitoring [9], [23].

III. Problem Identification and Future Directions

A. Solved Problems and Achievements

The integration of magnetic microwires into adhesive layers has successfully addressed several critical challenges in structural health monitoring [6]. A primary achievement is overcoming the invasiveness of traditional embedded sensors. While standard optical glass fibers (typically ~145 μm in overall diameter) have been shown to significantly reduce the quasi-static tensile shear strength of single-lap joints by increasing adhesive thickness and trapping air inclusions, amorphous glass-coated microwires (with a total diameter not exceeding 35 μm) can be embedded directly into the composite material without causing structural violations [9], [21]. The implementation of the enhanced induction method has also successfully enabled contactless diagnostics, allowing the measurement of tensile stress responses without the need for galvanic connections [7].

B. Open Problems and Current Limitations

Despite these advancements, several open problems remain before the technology can be universally deployed in real-world aerospace or civil infrastructure. One of the most pressing limitations is signal attenuation. Currently, the reliable detection of the microwire's switching field H_{SW} often requires the sensing and excitation coils to encircle the inspected beam or joint, which is highly impractical for large-scale, continuous structures such as aircraft wings or long concrete beams [21]. Signal attenuation at depths below a few millimeters in thick composite or metal adherends further restricts the sensor's effective range. Additionally, in conductive structures like carbon fiber reinforced polymers (CFRP), distinguishing the microwave or inductive signals originating from the ferromagnetic microwire inclusions from those generated by the carbon fibers themselves requires complex low-frequency modulating magnetic fields. Finally, interpreting multiaxial stress states and filtering out ambient magnetic noise in industrial environments poses significant signal processing challenges.

C. Feasible Issues and Proposed Future Directions

Addressing these limitations requires a multidisciplinary approach, and several issues are highly feasible to solve in the near term. To overcome the constraints of the encirclement detection method, future work must focus on proper signal filtering and amplification techniques. Enhancing the signal-to-noise ratio will allow the use of planar or vertically oriented coils placed on one side of the structure, rather than requiring full encirclement [21]. Another highly promising direction for further work involves the integration of advanced signal processing and machine learning algorithms. Machine learning can be utilized to interpret complex, non-linear magnetoelastic responses under 3D mechanical stress conditions [18]. This will help separate environmental noise from genuine structural degradation. Furthermore, experimental testing must transition from controlled, static single-lap joint coupons to real-time, dynamic monitoring of full-scale structural components subjected to prolonged hygrothermal aging. Investigating the long-term effects of moisture absorption and thermal cycling on the adhesion between the glass coating of the microwire and the polymer matrix will be crucial for guaranteeing the reliability of these sensors in harsh operational environments. Ultimately, optimizing the annealing processes (such as stress-annealing) to further enhance the giant magnetoimpedance (GMI) effect and domain wall dynamics will pave the way for even more sensitive, tunable composite materials for SHM [17].

IV. Conclusion

Investigating the issue of non-destructive evaluation of adhesive joints through embedded smart sensory networks represents an important front line in modern materials engineering. The integration of amorphous glass-coated magnetic microwires into critical adhesive joints brings a highly promising approach to structural monitoring. While thick optical sensors disrupt the continuity of polymer adhesives, thin composite microwires (up to 35 μm) significantly minimize any interfering effects on the structural integrity of the bonded joint.

This paper, by comprehensively analyzing the mechanics of lap joints, the chemistry of structural adhesives, the physics of ferromagnetic domains, and the metrological components of contactless induced signal transmission, highlights the strong potential and practical relevance of magnetic microwires for modern engineering. Utilizing these acquired insights and simulations provides a solid theoretical foundation for successfully demonstrating the functionality of the proposed sensory solution experimentally and for its systematic practical elaboration in future Structural Health Monitoring systems.

Acknowledgment

This research was supported by VEGA Grant No. 1/0350/24 from the Scientific Grant Agency of the Ministry of Education, Reasearch, Development and Youth of the Slovak Republic.

REFERENCES

[1] Kumar, S., Sivakumar, M. S. and Kalaichelvan, K. "Fatigue life prediction and damage assessment of adhesively bonded joints: A review". Polymer Testing. Vol. 93, p. 106926. ISSN 0142-9418. 2021. DOI: https://doi.org/10.1016/j.polymertesting.2020.106926

[2] Khalili, S. M. R., Jafarian, M. and Daghighi, S. "Enhancing Fatigue Life and Strength of Adhesively Bonded Composite Joints: A Comprehensive Review". Materials. Vol. 16, no. 19, p. 6468. ISSN 1996-1944. 2023. DOI: https://doi.org/10.3390/ma16196468

[3] Malekinejad, H.; Carbas, R.J.C.; Akhavan-Safar, A.; Marques, E.A.S.; Castro Sousa, F.; da Silva, L.F.M. Enhancing Fatigue Life and Strength of Adhesively Bonded Composite Joints: A Comprehensive Review. Materials 2023, *16*, 6468. https://doi.org/10.3390/ma16196468

[4] Kannan, S. "Acoustic-Emission Monitoring of Lap Joint Fatigue Cracks". Columbia: University of South Carolina. Dissertation thesis. 2023. Available from: https://scholarcommons.sc.edu/etd/7635

[5] Konka, Hari Prasad, Wahab, M. A. and Lian, Kun. "The effects of embedded piezoelectric fiber composite sensors on the structural integrity of glass-fiber–epoxy composite laminate". *Smart Materials and Structures*. Vol. 21, no. 1, p. 015016. ISSN 0964-1726. 2011. DOI: 10.1088/0964-1726/21/1/015016

[6] Zhukov, A., Talaat, A., Iriarte, M., Blanco, J. M. and Zhukova, V. "Development of glass-coated magnetic microwires for SHM sensors". Journal of Alloys and Compounds. Vol. 814, p. 152225. ISSN 0925-8388. 2020. DOI: https://doi.org/10.1016/j.jallcom.2019.152225

[7] Varga, R., Richter, K., Zhukova, V. and Zhukov, A. "Contactless stress monitoring using domain wall dynamics in magnetic microwires". *Scientific Reports*. Vol. 9, p. 13425. ISSN 2045-2322. 2019. DOI: https://doi.org/10.1038/s41598-019-49852-6

[8] Budhe, S., Banea, M. D., De Barros, S. and Da Silva, L. F. M. "An overview of adhesion and adhesively bonded joint analysis". International Journal of Adhesion and Adhesives. Vol. 106, p. 102813. ISSN 0143-7496. 2021. DOI: https://doi.org/10.1016/j.ijadhadh.2020.102813

[9] Varga, R., Richter, K., Sabol, R., Hudák, A., Šmelko, M. and Prasličkа, D. "Microwire-based sensors for structural health monitoring of composite materials". Sensors and Actuators A: Physical. Vol. 333, p. 113264. ISSN 0924-4247. 2022. DOI: https://doi.org/10.1016/j.sna.2021.113264

[10] Stuppiello, S., Marannano, G. and Zuccarello, B. "Analytical and numerical investigation of eccentricity effects in single-lap joints". *International Journal of Adhesion and Adhesives*. Vol. 128, p. 103550. ISSN 0143-7496. 2024. DOI: https://doi.org/10.1016/j.ijadhadh.2023.103550

[11] He, X., Liu, X., Zhang, G. and Wang, Y. "Numerical Analysis on Optimal Adhesive Thickness in CFRP Single-Lap Joints Considering Material Properties". Materials. Vol. 18, no. 11, p. 2423. ISSN 1996-1944. 2025. DOI: https://doi.org/10.3390/ma18112423

[12] Białoń, P., Smykla, J. and Muzia, G. "Comparative analysis of analytical and numerical models for single-lap adhesive joints". Materials. Vol. 16, no. 4, p. 1450. ISSN 1996-1944. 2023. DOI: https://doi.org/10.3390/ma16041450

[13] Galvez, P., Quesada, A., Martinez, M. A., Abenza, M. and Boada, M. J. L. "Effect of the adherend thickness and the overlap length on the peel stresses of single lap joints". The Journal of Adhesion. Vol. 95, no. 5-7, pp. 411-435. ISSN 0021-8464. 2019. DOI: https://doi.org/10.1080/00218464.2018.1511993

[14] Baykara, C., Ekrem, K. and Avci, A. "Characterization of Bending Strength in Similar and Dissimilar CFRP/Aluminum Single-Lap Adhesive Joints". Applied Sciences. Vol. 13, no. 23, p. 12879. ISSN 2076-3417. 2023. DOI: https://doi.org/10.3390/app132312879

[15] Anyfantis, K. N. "On the stress concentrations in bonded joints: A focus on the edges". Fatigue & Fracture of Engineering Materials & Structures. Vol. 45, no. 2, pp. 345-362. ISSN 8756-758X. 2022. DOI: https://doi.org/10.1111/ffe.13601

[16] Zhang, Y., Zhao, J., Zhou, A. and Ou, J. "Fatigue behavior of CFRP/steel bonded joints under cyclic tensile loading". Composite Structures. Vol. 280, p. 114886. ISSN 0263-8223. 2022. DOI: https://doi.org/10.1016/j.compstruct.2021.114886

[17] Kaur, M., Singh, J. and Sharma, S. "Recent progress in macromolecular network fatigue monitoring via embedded sensors". Progress in Polymer Science. Vol. 140, p. 101780. ISSN 0079-6700. 2024. DOI: https://doi.org/10.1016/j.progpolymsci.2023.101780

[18] Zhou, J., Xie, W. and Li, Y. "Machine learning-assisted stress interpretation for embedded magnetic sensors". Advanced Intelligent Systems. Vol. 5, no. 8, p. 2200450. ISSN 2640-4567. 2023. DOI: https://doi.org/10.1002/aisy.202200450

[19] Naghashzadeh, M., Ebrahimnezhad-Khaljiri, H. and Eslami-Farsani, R. "Challenges in self-sensing nanocomposites: From filler dispersion to signal reliability". Nano Materials Science. Vol. 6, no. 1, pp. 12-35. ISSN 2589-9651. 2024. DOI: https://doi.org/10.1016/j.nanoms.2023.05.002

[20] Hou, T., Sharples, S. D., Morris, E. and Hayes, S. A. "Embedded fiber Bragg grating sensors for internal strain monitoring of adhesive joints". Measurement. Vol. 151, p. 107125. ISSN 0263-2241. 2020. DOI: https://doi.org/10.1016/j.measurement.2019.107125

[21] Salajarvi, M., Klein, A. and Schmidt, R. "Optimization of planar pick-up coils for wireless SHM using magnetic microwires". Journal of Intelligent Material Systems and Structures. Vol. 36, no. 2, pp. 188-201. ISSN 1045-389X. 2025. DOI: https://doi.org/10.1177/1045389X241234567

[22] Zhukova, V., Iriarte, M., Talaat, A., Blanco, J. M. and Zhukov, A. "Tailoring of magnetic properties of glass-coated microwires for sensor applications". Journal of Magnetism and Magnetic Materials. Vol. 449, pp. 113-122. ISSN 0304-8853. 2018. DOI: https://doi.org/10.1016/j.jmmm.2017.10.021

[23] Richter, K., Sabol, R., Hudák, A., Zhukov, A. and Varga, R. "Stress sensitivity of the switching field in bistable magnetic microwires". IEEE Transactions on Magnetics. Vol. 59, no. 11, pp. 1-5. ISSN 0018-9464. 2023. DOI: https://doi.org/10.1109/TMAG.2023.3289001

Analysis of IoT Landscape in Education

[1]*Tomáš KORMANÍK (3rd year),*
Supervisor: [2]*Jaroslav PORUBÄN*

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]tomas.kormanik@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—This article focuses on the analysis of Internet of Things courses at other educational institutions. During our previous work, we determined that understanding the Internet of Things is often disputed. This statement is supported by the fact that various articles in this field are stating different, sometimes even contradictory, definitions of the Internet of Things. This realisation led us to our work published in a previous issue of SCYR, which raised several questions in regard to understanding the Internet of Things in an academic environment. Based on these findings, we are analysing numerous courses from other institutions with the goal of finding the state-of-the-art course and, if possible, determining the ideal Internet of Things curricula which aligns with its idea and definition and brings academic and industrial understanding to even ground.**

***Keywords*—Internet of Things, Education, Ambient Software, Higher Education**

I. Introduction

The paradigm of the Internet of Things (abbr. IoT) has been present in the scene for over 40 years and is constantly evolving. This evolution is enabled by technological advances in the fields of computer science and electronics in general. As we determined in our review [1], understanding of IoT in the past was initially correct but was limited by current capabilities of technology and infrastructure. In the past 15 years, the popularity of IoT has increased, since current technologies unlock the potential of this paradigm. In many cases, IoT implementations have become parts of our everyday lives, often without us even realising it. In our aforementioned conference paper, we presented Apple AirTag as a easy-to-understand example for students, since it allows us to easily explain benefits but also possible drawbacks or even risks of IoT thanks to events that actually happened since its appearance on the consumer market. There are numerous other examples which align with the IoT paradigm, such as *Google Maps*, which collects anonymised data from each of its users in order to provide them with the most ideal route estimates.

Furthermore, this concise conference overview of IoT indicated to us that there is a significant gap between academic education and industrial implementation of IoT. While some gaps are natural due to patents or greater investment, many of these solutions have been publicly available for several years without being incorporated into education. Since their introduction, both *Raspberry Pi* and *Arduino* have become popular in the education of IoT; however, we felt that the majority of the courses teaching with them are only labeled as IoT while actually teaching Smart Devices.

II. Current background and progress

During our work in the previous issue of SCYR[2], we presented our ongoing analysis of the current state of IoT education based on the history of IoT and published articles. In addition, we showcased a hypothetical example that could serve as a demonstration of IoT capabilities for students. Thanks to the relevant feedback during the conference, we have begun the analysis of articles presenting IoT courses or education practices since their introduction to the educational process. After 20 analyzed courses we were able to clearly define standard structure of IoT course as well as distinct features of the course for different study focuses (electronics, informatics, software engineering). Results of our analysis were part of our next publication[1]. From a theoretical perspective, we have identified works by Rajora et al. [3] and Natalia Silvis-Cividjian [4] as theoretically correct and worth consideration when designing whole IoT-orientated curricula or independent courses.

These findings have made obvious to us that there is a significant gap in understanding of IoT between academic and industrial fields. Companies such as *Cisco* are providing various new use cases of IoT, which is natural, since their products are driving the IoT networks. When we looked at courses and curricula pertaining to IoT, even if they stated or indicated the correct definition, their laboratory assignments or semestral projects often didn't align with it at all. An example scenario that is really common and aligns with the IoT definition is intelligent waste management. In this area, we can find numerous articles and solutions that go to different depths. Wang et al. [5] present a solution which uses CNN[1] for identification of waste type, which then leads to correct waste processor arrival when needed. While this solution is relatively recent, it is only different in CNN[1] implementation compared to the solution by Kumar et al. [6], which is older and focuses more in detail on route planning. This paper also demonstrates how possibilities of IoT solutions and their possible applications have widened with the popularisation of artificial intelligence.

III. Future research

Our current focus is to review recent publications from the start of the year 2024 until the end of the year 2025 which are related to education in IoT. This will provide us with clear information about current trends in the field while exposing general trending topics and innovative approaches. So far, we

[1]Convolutional Neural Network

have reviewed approximately 190 articles and publications. During our review, we have determined that each of the publications can be assigned to one of these 7 categories:

1) **Not applicable** - publications contain IoT only as a keyword without any relation to it, or publications only vaguely mention possible use in IoT applications without any details. Work by Fridman et al. [7] presents their augmented reality solution for the production line, but the clear link to IoT or description of the the production line is missing. So far, these account to 7,69% of results in one scientific database.
2) **Relevant conceptual or theoretical work** - publications which are focusing on underlying technologies of IoT such as protocols, patents, hardware, and training or adjusting Artificial Intelligence models. A great example is the work by Fabio Petrillo [8], which reviews whether the Rust platform is suitable for IoT solutions. So far, these the majority of results in one scientific database, with 41,53% share.
3) **Single device with internal communication** - implementation or used example utilises a single device or Single Board Computer (abbr. SBC) which communicates internally with its sensors and services. Zhumaniyaz Mamatnabiyev presents their use of the IoTXplorBot robot, which allows students to create a device which physically interacts with the world. These are only 4,62% of results in one scientific database.
4) **Single device with external communication** - implementation or used example utilises a single device or SBC which communicates with anything in the network (for example, a database, API, website and backup server). A wearable device collecting physiological and biomechanical data, which is sent to the cloud for data processing, evaluation and response, is presented by Normatova et al.[9]. These solutions are almost 3 times more popular than singular ones, with 13,85% publications falling into this category.
5) **Multiple devices with isolated communication** - the solution utilises multiple devices and SBCs along with their sensors in order to exchange data through their private or non-public network. Paliy et al. [10] are presenting their solution, which utilises numerous sensors present at their local university buildings, which provide data that enables precise management of heating and cooling at the premises. 12,31% of publications contain solutions falling into this category.
6) **Multiple devices with external communication** - solution utilises multiple devices and SBCs which are utilising public/private databases, endpoints, and cloud services to store and process their data. Satyajeet Shinge and Urmila Shrawankar [11] write about their smart university solution, which is similar to the one above, but it utilises the cloud for analysis and publishing of data. This group is so far the most popular in publications with actual solutions or examples in them - with 18,46% share.

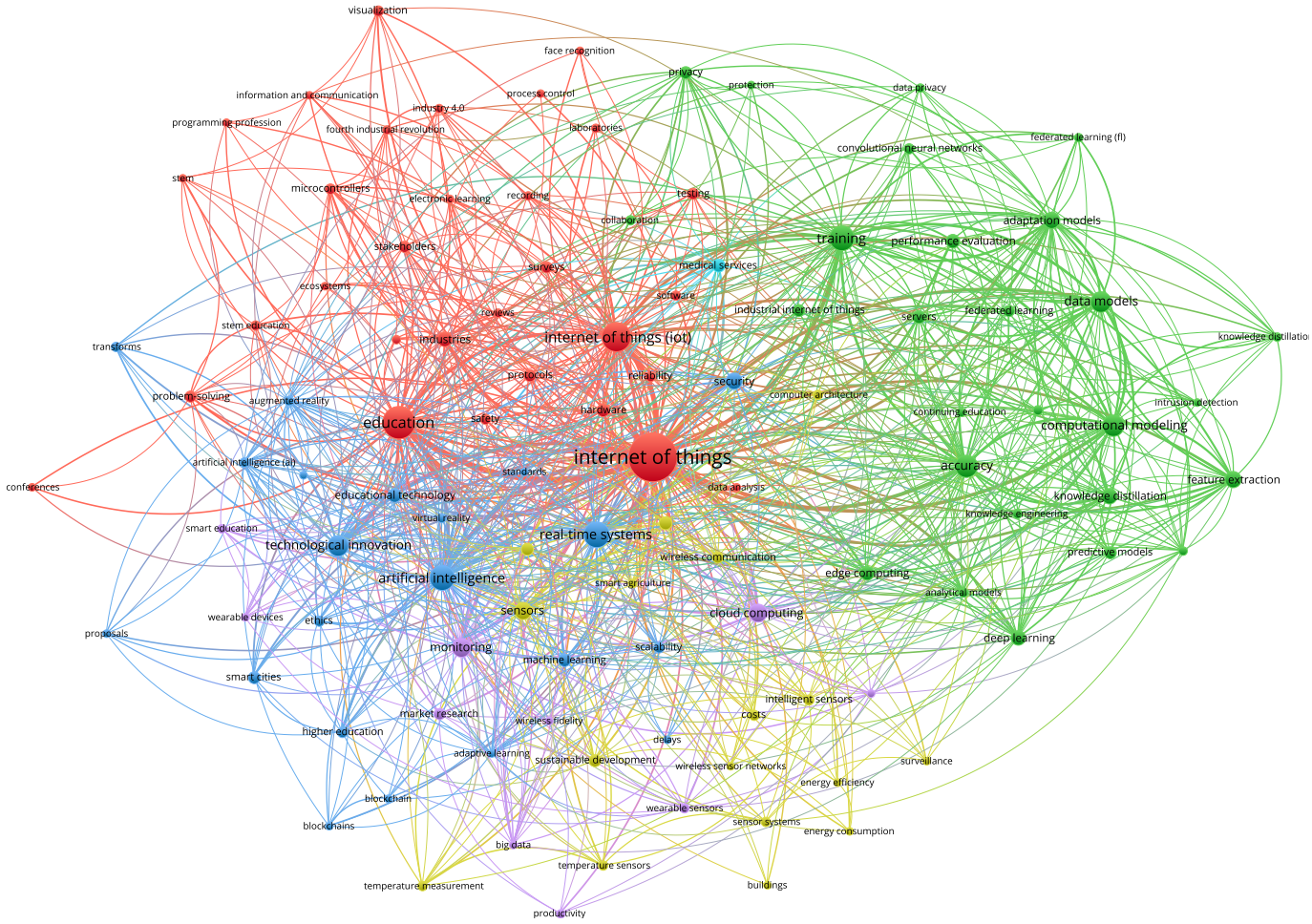


Fig. 1. Cluster diagram of keyword relations from *IEEE Xplore* research database query results generated using VOSviewer

7) **Network of mutually beneficial devices** - network of devices which independently communicate with each other in order to improve functionality of themselves and others. Zuniga et al.[12] created an outline for the ideal structure of a Massive Online Open Course (abbr. MOOC), which can be used as inspiration for general IoT courses as well, since their suggestions align with the paradigm of IoT. Sadly, this solution was only a singular occurence, so its share is less than 1%.

All of the examples mentioned above are from the *IEEE Xplore* research database, since we have finished analysing data from it first. A great way to visually look at clusters of related keywords and topics is by using a cluster diagram (Fig. 1), where each cluster is identified by a different colour. In our case, we utilised the VOSviewer[2] software.

While conference papers and articles usually present example solutions utilised in IoT courses, we are more interested in actual courses themselves – mainly their curricula, theoretical prerequisites, lecture contents, assignments and evaluation processes. So far, data indicates that IoT courses are taught in the 2nd or 3rd year of a bachelor's programme and rarely in the 1st year of a master's programme. Out of the courses referenced in publications so far, Zuniga et al.[12] have made their MOOC publicly accessible at https://iot.mooc.fi/ for non-industrial use. Naturally, most of the courses don't have new articles about them published regularly, so finding them requires searching through webpages of individual universities and their curricula. The most detailed variant found using conventional methods is taking place at Czech Technical University in Prague and is available at URL https://cw.fel.cvut.cz/wiki/courses/b0b37nsi.

IV. Conclusion

IoT courses and education overall are evolving but, in our opinion, significantly lagging behind the industrial state-of-the-art. While electronics-orientated study programmes should look at IoT hardware in more detail, in our opinion, the current over-reliance on SBCs has become present in computer science and software engineering study programmes as well.

The paradigm of IoT will continue to grow since technologies like AI are being adjusted or compressed into smaller runtimes or partial models, which enable them to run on devices with less computing power or by using ensemble learning, as presented by Wu et al.[13]. Another popular approach is the use of fog computing, which distributes the workload between the IoT device itself, some servers in the local network, or a high-performance cloud service. These solutions all together will allow not only for more complex processing possibilities at the IoT devices but also enable strengthening of IoT security, as, for example, Muhammad et al.[14] demonstrate with their solution against SQL injections.

V. AI Acknowledgement

I acknowledge the use of *QuillBot*[3] to verify spelling and grammar in this article. Furthermore, *Google Gemini*[4] was used as a secondary tool after conventional translation tools in order to verify contextual understanding of written text in languages different than English, which were encountered and analysed during our research.

[2]VOSviewer webiste - https://www.vosviewer.com/
[3]QuillBot URL - https://quillbot.com/
[4]Google Gemini URL - https://gemini.google.com/

VI. Funding Acknowledgement

This and all related works and publications were supported by following parties or projects in last year:

- Project KEGA No. 061TUKE-4/2025 Building Bridges between University and High School ICT Education.
- The Slovak Research and Development Agency under the Contract no. APVV-23-0408.

References

[1] T. Kormaník and J. Porubän, "Rethinking IoT Education: Is the Concept Truly Grasped?" in *6th International Computer Programming Education Conference (ICPEC 2025)*, ser. Open Access Series in Informatics (OASIcs), R. Queirós, M. Pinto, F. Portela, and A. Simões, Eds., vol. 133. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, pp. 11:1–11:8. [Online]. Available: https://drops.dagstuhl.de/entities/document/10.4230/OASIcs.ICPEC.2025.11

[2] T. Kormaník, "Towards iot education evolution," in *SCYR 2025: 25th Scientific Conference of Young Researchers Proceedings from Conference*. Košice, Slovak Republic: Faculty of Electrical Engineering and Informatics, Technical University of Košice, 2025, pp. 18–20.

[3] R. Rajora, A. Rajora, B. Sharma, P. Aggarwal, and S. Thapliyal, "The integration of iot in learning environments: Assessing the impact of educational empowerment and addressing its challenges," in *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 2024, pp. 1–5.

[4] N. Silvis-Cividjian, "Teaching internet of things (iot) literacy: A systems engineering approach," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 2019, pp. 50–61.

[5] C. Wang, J. Qin, C. Qu, X. Ran, C. Liu, and B. Chen, "A smart municipal waste management system based on deep-learning and internet of things," *Waste Management*, vol. 135, pp. 20–29, 2021.

[6] S. V. Kumar, T. S. Kumaran, A. K. Kumar, and M. Mathapati, "Smart garbage monitoring and clearance system using internet of things," in *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, 2017, pp. 184–189.

[7] S. Fridman, S. Berenson, and N. Unkelos-Shpigel, "Production line augmented reality application," in *2025 IEEE 33rd International Requirements Engineering Conference (RE)*, 2025, pp. 580–581.

[8] F. Petrillo, "Should we use rust platform in our iot applications? a multivocal review," in *2025 IEEE/ACM 7th International Workshop on Software Engineering Research Practices for the IoT (SERP4IoT)*, 2025, pp. 24–31.

[9] A. Normatova, A. Hamedshnain, M. Hameed, A. M. Ahmatkhonovich, and P. Manisha Prashant, "Wearable tech and iot for the future of fitness and physical education training system," in *2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES)*, 2025, pp. 1–7.

[10] S. Paliy, V. Druzhynin, O. Kuchanskyi, M. Gladka, Y. Hozak, and T. Honcharenko, "Iot technology for energy saving in educational buildings by accounting for human body heat," in *2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST)*, 2025, pp. 1–6.

[11] S. Shinge and U. Shrawankar, "Empowering education industries with a cloud based iot framework," in *2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0*, 2025, pp. 1–5.

[12] A. Zuniga, N. Thi Nguyen, M. Tompuri, H. Nygren, and P. Nurmi, "A tale of many iots: A modular constructivist course design for internet of things education," *IEEE Transactions on Education*, vol. 68, no. 6, pp. 516–527, 2025.

[13] S. Wu, Y. Li, W. Wang, J. Guo, W. Fan, Q. Liu, W. Jia, S. Yu, J. Cao, and T. Wang, "Enhancing collaborative inference on heterogeneous edge devices via adaptive ensemble knowledge distillation," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 10, pp. 3564–3579, 2025.

[14] H. Y. Muhammad, V. Suryani, and A. G. Putrada, "Knowledge distillation for sql injection detection in iot edge dashboards using bilstm," in *2025 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2025, pp. 206–211.

Transparent Intensional Logic from a Programming Language Perspective

[1]*Samuel NOVOTNÝ (3rd year),*
Supervisor: [2]*William STEINGARTNER*

Department of Computers and Informatics, FEI, Technical University of Košice, Slovak Republic

[1]samuel.novotny@tuke.sk, [2]william.steingartner@tuke.sk

***Abstract*—Although the procedural nature of Transparent Intensional Logic might suggest that the system is naturally predisposed to practical implementation, this aspect has received relatively little attention, since its development has been driven primarily by applications in the logical analysis of natural language. Therefore, in this paper, we summarize our recent work examining this logical framework from a programming language perspective. The resulting output of this effort includes a refinement of its type system, the implementation of its functional framework in Haskell, and the introduction of logical spacetime as its explicit extensional model.**

***Keywords*—constructions, extensional model, Haskell, Transparent Intensional Logic, type incoherence, type system**

I. Introduction

Tichý's Transparent Intensional Logic (TIL) [1] represents one of the principal fine-grained methodologies within the field of logical analysis of natural language (LANL) [2]. Given its primary application domain, this logical framework has been cultivated almost exclusively within the community of logicians and analytic philosophers, who seek to provide a meaning (Fregean sense) of natural language expressions in a mathematically precise way. In this traditional setting, the formal apparatus of TIL, hyperintensional modal-temporal partial typed λ-calculus, has primarily served as an instrument for LANL performed manually by trained logicians. As a consequence, certain aspects of the system have not been examined with the degree of formal rigor commonly associated with programming languages, of which TIL is, in a precise technical sense, an instance.

Therefore, in the present paper, we adopted a different perspective and examine TIL from the viewpoint of programming language theory. Such a shift naturally raises questions concerning the explicit formulation of its type system, the possibility of systematic type derivation, and the feasibility of computational realization.

The aim of this paper is to summarize recent work in which TIL is reconsidered as a programming-language-like logical system. This development proceeds in three interrelated steps: (Section III-A) formal refinement of the type system [3], (Section III-B) functional implementation in Haskell [4], and (Section III-C) construction of logical spacetime as an explicit extensional model [5]. Together, these steps demonstrate that TIL may be interpreted not only as a semantic framework, but also as a computationally realizable strongly typed logical system.

II. Background: Transparent Intensional Logic

TIL is a hyperintensional variant of a typed λ-calculus equipped with procedural semantics. Thus, TIL λ-terms are furnished with algorithmically structured abstract procedures called constructions that produce functions-in-extension, i.e., set-theoretical mappings or even further constructions. In contrast to the simply typed λ-calculus, there are six kinds of construction. Two of them are atomic constructions, Trivialisation and variable, which supply objects on which molecular constructions can operate.

Variables $x, y, p, q, \ldots$ v-produce objects of their respective ranges dependently on a valuation v. *Trivialisation* of an object X, ${}^0X$ in symbols, is a one-step procedure that simply produces the object without any change; ${}^0X$ is not unlike a constant in a standard λ-calculus.

Molecular constructions are Closure and Composition. A *Closure* $[\lambda x_1 \ldots x_n C]$ is a procedure that v-produces a functions f with values produced by C and arguments obtained by abstracting over the values of variables $x_1, \ldots, x_n$. *Composition* $[C_0 C_1 \ldots C_n]$ v-produces the value (if any) of the function v-produced by C_0 at the n-tuple of arguments v-produced by $C_1, \ldots, C_n$.

In addition to the above constructions, there are two special ones, namely Execution and Double Execution. Where C is a construction, the *Execution* ${}^1C$ v-produces what C v-produces (if anything). Hence, ${}^1C$ is like C, only that the order of ${}^1C$ is $n+1$ where n is the order of C; ${}^1C$ in standard TIL is almost of no effect.On the other hand, *Double Execution* is a much-needed construction. Where C is a construction, the Double Execution ${}^2C$, v-produces what is v-produced by the product of C (if anything). Hence, where C is a construction, Trivialisation ${}^0C$ produces C thus introducing a hyperintensional context in which C and all its sub-constructions occur displayed (as arguments of other functions produced within the super-procedure in which C occurs). Yet, sometimes we need to cancel the effect of Trivialisation and turn the mode of occurrence of C to executed. To this end, Double Execution is applied, as there is a universally valid rule of 20-elimination: ${}^{20}C = C$.

The ontology of TIL is organized in a ramified hierarchy of types combining Church's simple theory of types, extended by the atomic types of time moments τ and possible worlds ω to capture temporal and intensional variability, with Russell's ramified theory of types, which enforces stratification of procedural entities (types $*_n$) and thereby yields hyperintensionality.

Traditionally, typing in TIL is expressed using two relations:

$$X/\alpha, \qquad C \to \alpha. \tag{1, 2}$$

The relation (1) states that the object X is a member of type α, whereas the relation (2) states that the construction C constructs (if anything) an object of type α.

In the standard methodological practice of TIL, type checking constitutes the third phase of logical analysis. However, the interaction between the relations $/$ and $\to$ has been treated largely semi-formally, without an explicitly formulated derivation system or a complete set of typing rules. While this semi-formal status was sufficient for manual logical analysis, it leaves open questions concerning formal rigor and computational interpretation.

III. Results

Viewing TIL through the lens of programming language theory shifts attention from its application in a specific LANL domain to the formal system of TIL itself.

A. Type-Theoretic Refinement of TIL

To obtain a derivation-based formulation compatible with standard approaches to typed λ-calculus [6], typing in TIL must be expressed relative to explicit contexts.

Accordingly, we introduced a typing relation (3) in the sequent form where Γ is the construction context assigning types to free procedural variables, Δ is the object context assigning types to non-procedural entities, X is the typed entity, α is its membership type, and β is the type of the object constructed by X if X is construction.

$$\Gamma \mid \Delta \vdash X/\alpha \to \beta \tag{3}$$

In the course of formalizing the typing discipline, we introduced explicit derivation rules for all kinds of TIL constructions, thereby reconstructing the semi-formal typing discipline as a complete rule-based formal system. The formulation was guided by the Prolog implementation of the type checking algorithm described by Duží and Fait [7]. While these rules appear formally coherent, a detailed analysis [3] revealed a structural problem in their interaction. In particular, the issue arises from the combined application of the following three rules:

$$\frac{\Gamma \mid \Delta \vdash X/\alpha}{\Gamma \mid \Delta \vdash {}^0X \to \alpha}\ \text{(T-triv)} \qquad \frac{\Gamma \mid \Delta \vdash C \to \alpha}{\Gamma \mid \Delta \vdash C/*_n}\ \text{(T-prog)}$$

$$\frac{\Gamma \mid \Delta \vdash X \to *_n}{\Gamma \mid \Delta \vdash {}^2X \to \alpha}\ \text{(T-2exec)}$$

Rule (T-prog) causes a loss of type information about the object constructed by C, since the type $*_n$ carries only order information but not the type of the constructed object. Conversely, rule (T-2exec) introduces object-type information in its conclusion without sufficient justification at the derivation level.

As a consequence, the original system admits constructions that are type-theoretically incoherent yet remain undetected both by standard type checking and by the type inference procedure. In particular, the constraint set generated during type inference may remain satisfiable despite the internal inconsistency of the construction, leading to successful unification and erroneous classification of the term as well-typed.

The source of this incoherence lies in the definition of construction types. In the original hierarchy, the type $*_n$ encodes only the order of a construction, not the type of the object it constructs. To eliminate this information loss, construction types are redefined so that they explicitly carry the type of the constructed object. Instead of $*_n$, we introduced types (4) denoting constructions of order n constructing objects of type α.

$$(\alpha *_n) \tag{4}$$

Under this refinement, the problematic rules are replaced by the following rules:

$$\frac{\Gamma \mid \Delta \vdash C \to \alpha}{\Gamma \mid \Delta \vdash C/(\alpha *_n)}\ \text{(T-prog}')$$

$$\frac{\Gamma \mid \Delta \vdash X \to (\alpha *_n)}{\Gamma \mid \Delta \vdash {}^2X \to \alpha}\ \text{(T-2exec}')$$

The modified typing discipline preserves object-type information throughout the derivation process. As shown in [3], constructions for which the original system admitted successful type derivations, despite their internal type-theoretical incoherence, now give rise to unsatisfiable constraint sets during type inference. Consequently, unification fails and the incoherence is correctly detected.

This refinement establishes TIL as a derivation-based strongly typed formal system whose typing discipline aligns with standards commonly associated with programming language theory, while preserving its hyperintensional and ramified structure.

B. Implementation of TIL in Haskell

Once the typing discipline was formalized, TIL became ready for systematic computational realization. Although the TIL-Script language was previously developed as a computational variant of TIL [7], it was designed as a standalone language with its own syntax and interpreter. In contrast, our approach was motivated by a different idea: instead of constructing a new programming language from scratch, we investigated whether an existing strongly typed functional language could serve as a host environment for TIL.

The key insight is that Haskell already provides a rich universe of non-procedural objects together with a statically typed higher-order functional framework. Rather than implementing TIL as an independent language, we define a TIL-framework within Haskell, in which TIL constructions are simulated by so-called *constructional functions*.

The correspondence is summarized conceptually in Table I.

TABLE I: Formal Correspondence between TIL and Its Haskell Framework

TIL	TIL-framework in Haskell
construction	constructional function
construction execution	application of constructional function to valuation
type of construction $(\alpha *_n)$	type `Star -> alpha`

Under the redefined notion of construction type $(\alpha *_n)$, a construction can be viewed operationally as a function

from valuation to the type α of its product. In the Haskell framework, this is represented by the type (5):

```
Star -> alpha,                    (5)
```

where `Star` represents valuation, i.e., the procedural context relative to which constructions are evaluated.

To model valuation explicitly, we introduce the structured data type `Star`, whose example definition is given in Code I.

```haskell
data Star = Star {
  x :: Tau,
  -- further variables
}
```

CODE I: Implementation of valuation in Haskell

Each free variable is implemented as a projection from the valuation object `Star`. In this sense, free variables are functions extracting their assigned value from the procedural context.

The primitive operators of the TIL language, Trivialization $^0(\ )$, Composition, Closure, Single Execution $^1(\ )$, and Double Execution $^2(\ )$, are implemented in the Haskell framework by the functions `triv`, `comp`, `clos`, `exec1`, and `exec2`, respectively. This implementation does not use the standard TIL or TIL-Script notation, but is inspired by the notation from Tichý's original work [8], which better reflects the simulation aspect of the framework.

These functions correspond to the primitive operators of TIL and do not represent constructions themselves. A construction, more precisely, a constructional function, arises only after supplying the required arguments to the corresponding primitive.

Their Haskell type signatures are presented in Code II.

```haskell
triv  :: alpha -> Star -> alpha

comp  :: (Star -> alpha -> beta)
         -> (Star -> alpha)
         -> Star -> beta

clos  :: ((Star -> alpha) -> Star -> beta)
         -> Star -> (alpha -> beta)

exec1 :: (Star -> alpha)
         -> Star -> alpha

exec2 :: (Star -> Star -> alpha)
         -> Star -> alpha
```

CODE II: Type signatures of Haskell implementations of TIL primitives

In this framework, Haskell provides non-procedural objects (numbers, functions, data types) directly, while TIL primitives simulate the procedural layer. The result is not a reimplementation of TIL as an independent programming language, but a typed TIL-framework embedded within a strongly typed functional host language.

An important consequence of this embedding is the emergence of a dual-layer type discipline. Internal TIL typing, formalized in the previous section, is complemented by Haskell's static type system, which enforces additional consistency at compile time. Consequently, TIL may be viewed not merely as a semantic methodology, but as a domain-specific strongly typed functional language with hyperintensional semantics.

C. Construction of an Extensional Model of TIL

The procedural semantics of TIL enables constructions to produce intensions (functions from possible worlds). For instance, a construction of a proposition yields a function of type $((o\tau)\omega)$ mapping possible worlds and time moments to truth values. For logical analysis of natural language, such an intensional result is fully sufficient.

From a computational perspective, however, it is often necessary to determine the extensional value of a construction at a specific possible world w and time moment t representing the state of the system. In applications such as verification and model checking, one is not merely interested in the intension itself, but for example, in whether a given proposition is true at a particular state of the system.

To provide a formally defined evaluation environment, we introduced the notion of *logical spacetime* – the single unique and real underlying structure of the extensional model of Transparent Intensional Logic in [5] as the ordered structure

$$\mathcal{LST} = (S, \prec),$$

where S is the set of all world-time points (w-moments) and $\prec$ is a binary accessibility relation on S.

Unlike in Kripke-style semantics, where the accessibility relation is introduced as a primitive component of the frame, the relation $\prec$ in logical spacetime is induced from the naturally existing strict linear ordering on the set of time moments.

Acknowledgment

This research was supported by the Cultural and Educational Grant Agency (KEGA) under project No. 052TUKE-4/2025, "Modern Approaches in Educating IT Professionals in the Field of Type Theory" and by KEGA project 028TUKE-4/2026 – "Innovative approaches to teaching and researching formal methods in the context of modern software engineering", granted by the Cultural and Education Grant Agency of the Slovak Ministry of Education.

References

[1] M. Duží, B. Jespersen, and P. Materna, *Procedural semantics for hyperintensional logic: Foundations and applications of transparent intensional logic*. Springer Science & Business Media, 2010, vol. 17. [Online]. Available: https://doi.org/10.1007/978-90-481-8812-3

[2] M. Duží and P. Materna, "Logical form," *Essays on the Foundations of Mathematics and Logic*, vol. 1, pp. 115–153, 2005.

[3] S. Novotný, M. Duží, and W. Steingartner, "Implementation of transparent intensional logic framework in haskell," in *Information Modelling and Knowledge Bases XXXVII*, ser. Frontiers in Artificial Intelligence and Applications, M. Menšík, H. Jaakkola, M. Tropmann-Frick, Y. Kiyoki, V. Sornlertlamvanich, and N. Yoshida, Eds., vol. 418. Amsterdam, Netherlands: IOS Press, 2026, pp. 388–398. [Online]. Available: https://doi.org/10.3233/FAIA251739

[4] S. Novotný, M. Duží, W. Steingartner, and J. Perháč, "Towards resolving type incoherence in transparent intensional logic," in *Information Modelling and Knowledge Bases XXXVII*, ser. Frontiers in Artificial Intelligence and Applications, M. Menšík, H. Jaakkola, M. Tropmann-Frick, Y. Kiyoki, V. Sornlertlamvanich, and N. Yoshida, Eds., vol. 418. Amsterdam, Netherlands: IOS Press, 2026, pp. 335–348. [Online]. Available: https://doi.org/10.3233/FAIA251735

[5] S. Novotný, W. Steingartner, and V. Novitzká, "On structure of tichý-inspired logical spacetime," *Journal of Applied Mathematics & Computational Mechanics*, vol. 24, no. 4, 2025. [Online]. Available: https://doi.org/10.17512/jamcm.2025.4.04

[6] B. C. Pierce, *Types and programming languages*. MIT press, 2002.

[7] M. Duží and M. Fait, "Type checking algorithm for the til-script language," *Frontiers in Artificial Intelligence and Applications*, vol. 312, pp. 219–36, 2019.

[8] P. Tichy, "Constructions," *Philosophy of Science*, vol. 53, no. 4, pp. 514–534, 1986.

Shadow Layer Protection: A Software-Defined Superstructure for Resilient Smart Grid Infrastructure

[1]*Marek Bobček (3rd year)*
Supervisor: [2]Zsolt Čonka

[1,2]Dept. of Electric Power Engineering, FEI, Technical University of Košice, Slovak Republic

[1]marek.bobcek@tuke.sk, [2]zsolt.conka@tuke.sk

***Abstract*—Modern power grids necessitate adaptive protection frameworks due to the increasing penetration of volatile renewable energy sources. Traditional hardware-based relays often rely on fixed threshold settings, which limits their operational flexibility under dynamic grid conditions. This paper proposes a hybrid multi-layer architecture featuring a Shadow Layer a software-defined virtual protection superstructure that operates in parallel with established digital relays. This shadow system leverages high-speed synchronized measurements from the AD7606 16-bit converter and Phasor Measurement Units (PMUs) to enhance real-time situational awareness. By integrating intelligent signal processing, including autonomous AC/DC mode detection via peak-to-peak analysis and precise True RMS calculations, the superstructure provides a critical layer of diagnostic granularity. Experimental validation involving 1,000 test samples confirms that the Shadow Layer achieves an ultra-low internal detection latency of 1 ms, while maintaining a total trip command time of 64.2 ms that is highly competitive with industrial benchmarks from SEL, ABB, and Siemens. This framework introduces a 2-out-of-5 validation logic to ensure transient immunity. Ultimately, the Shadow Layer serves as a sophisticated intelligence augmentation for existing legacy infrastructure, significantly boosting smart grid reliability without necessitating immediate hardware replacement.**

***Keywords*—Virtual Protection, Shadow Layer, Smart Grid, AD7606, True RMS, Fault Detection, Hybrid Protection Architecture.**

I. Introduction and motivation

The global transition toward decentralized energy systems and the massive integration of Renewable Energy Sources (RES) have fundamentally altered the dynamic behaviour of electrical power systems. Traditional protection schemes, which have served as the backbone of grid stability for decades, are increasingly challenged by bidirectional power flows and reduced fault levels characteristic of inverter-based resources. These legacy systems typically rely on fixed threshold settings and deterministic logic, which often lack the necessary sensitivity or response speed required for the rapid transients occurring in modern Smart Grids. The motivation for this research stems from the urgent need to bridge the gap between traditional hardware reliability and high-speed digital adaptability. By leveraging Wide Area Monitoring Systems (WAMS) and high-resolution synchronized data, it is possible to create a virtualized protection environment. However, a total replacement of physical hardware remains unfeasible due to cybersecurity concerns and the requirement for hardware determinism. Therefore, this work explores the implementation of a hybrid framework where a Shadow Layer serves as an intelligent parallel superstructure, enhancing fault detection capabilities while maintaining the safety net provided by traditional digital relays [1-8].

II. Initial Stage of the Research Tasks

The second year marked a transition toward the optimization of the communication architecture and physical hardware integration. A significant bottleneck identified in traditional WAMS setups was the latency introduced by intermediate processing units, such as Real-Time Automation Controllers (RTAC). This phase of the research, referred to as "Stage 2," involved streamlining data flow directly from the AD7606 high-speed sensing unit to virtual protection servers to minimize reaction times. Furthermore, advanced algorithmic validation protocols were developed to ensure the integrity of incoming data, confirming that the protection logic is not falsely triggered by measurement noise or communication dropouts a critical requirement for the reliability of the Shadow Layer superstructure [1-8].

III. Current Research Phase: Development of the Shadow Layer Framework

The current phase of the research focuses on the transition from wide-area theoretical models to a localized, high-speed Shadow Layer execution framework. Building on established foundations regarding data validation in Wide Area Measurement Systems (WAMS) and the possibilities of virtual protection, the research has moved toward the physical implementation of a soft-computing superstructure. This phase addresses the critical need for a protection layer that can process raw high-resolution signals in parallel with traditional digital relays without introducing the latencies typical of conventional Wide Area Measurement Protection and Control (WAMPAC) architectures. The primary objective is the development of a real-time monitoring and protection environment that remains agnostic to the underlying primary hardware, thereby functioning as a secondary intelligence layer. A significant component of this stage is the integration of high-resolution sensing hardware with a Python-based processing engine to ensure deterministic reaction times. This

approach allows for the application of adaptive signal characterization—specifically distinguishing between AC and DC operating modes—which provides the system with the flexibility to protect diverse grid components, from standard distribution lines to inverter-based renewable resources. This "Shadow Layer" is designed to validate fault conditions through multi-stage logic, such as the 2-out-of-5 validation protocol, providing a robust backup to legacy infrastructure while offering superior diagnostic granularity [1-8].

IV. Model Architecture

The core of the system is a high-performance Python-based HMI (Human-Machine Interface) and controller designed for real-time monitoring and protection in electrical distribution systems. The architecture is divided into three primary layers: Data Acquisition, Signal Processing, and Protection Logic. The Model Architecture is designed to function as an intelligent monitoring and decision-making node positioned as a superstructure above the physical hardware layer.

A. Data Acquisition Layer (DAQ)

The system utilizes a 16-bit AD7606 converter, which enables simultaneous sampling across eight independent channels. This specific hardware configuration ensures zero phase shift between measurements, which is a critical parameter for ensuring high-precision analysis within the Shadow Layer framework. Data is streamed via a high-speed serial interface at a baud rate of 921,600, with dedicated thread handling to maintain deterministic throughput.

1.) Communication Protocol: Data is streamed via a high-speed serial interface (UART) at a baud rate of 921,600.
2.) Buffer Management: To minimize latency, the SerialReader class operates in a dedicated thread, employing large RX/TX buffers to prevent data loss during high-frequency sampling.
3.) Raw Data Conversion: Incoming 16-bit signed integers are converted to voltage values using a pre-calculated coefficient: $V=raw\times(5.0/32768.0)$.

B. Intelligent Signal Processing Layer

This layer is responsible for characterizing the electrical signal and extracting meaningful parameters for the HMI and protection algorithms.

1.) Mode Detection: The system performs autonomous AC/DC mode detection through Peak-to-Peak analysis within a sliding window of 100 samples.
2.) True RMS Calculation: For identified AC signals, a True RMS algorithm is implemented, providing significantly higher accuracy for non-linear loads and distorted waveforms compared to standard Mean Absolute Value (MAV) industrial rectifiers.
3.) Noise Suppression: A software-defined noise gate with a threshold of 0.03 V is utilized to maintain stable "IDLE" states and prevent nuisance operations caused by low-level electromagnetic interference.

C. Protection Logic and HMI

The protection logic executes safety protocols and provides real-time visual feedback.

1.) Multi-Stage Validation (2-out-of-5): To ensure maximum reliability and immunity against transient disturbances, the architecture employs a "2-out-of-5" fault history protocol.
2.) Trip Execution: A hardware "TRIP" command is initiated only when the signal exceeds defined limits in at least two out of five consecutive evaluation windows, effectively filtering out momentary spikes or measurement errors.
3.) Adaptive Parameters: Each channel supports independent configuration for Under-Voltage (UV), Over-Voltage (OV), and specific time delays, enabling seamless coordination with existing legacy protection devices in the grid.

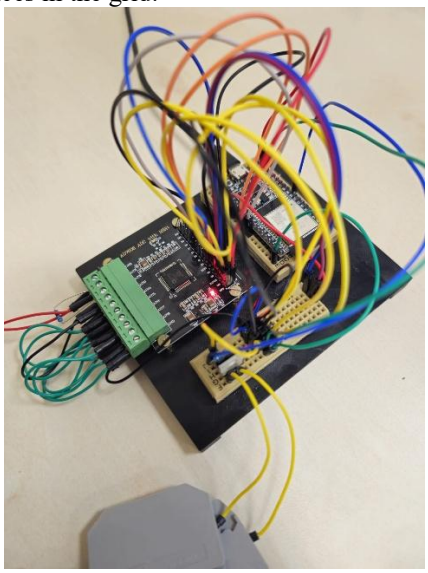


Fig. 1. The first system prototype for high-speed real-time data acquisition and processing, designed as the hardware foundation for the Shadow Layer virtual protection superstructure. - The prototype comprises a 16-bit AD7606 analog-to-digital converter for simultaneous 8-channel sampling with zero phase shift, a Python-based processing engine for deterministic signal analysis and True RMS calculation, and a high-speed UART interface (921,600 baud) with dedicated threading to minimize latency. The system integrates PMUs for WAMS synchronization and a PyQt6-based HMI for real-time visualization and control.

V. HMI Design and Implementation

The HMI (Human-Machine Interface) serves as the critical visualization and configuration layer of the ADCSYS platform, developed using the PyQt6 framework for high-performance graphical rendering. The interface is engineered to provide real-time situational awareness while maintaining a low computational footprint to ensure the timing accuracy of the underlying protection logic [6-8].

A. Real-Time Monitoring Interface

The primary monitoring window is designed for maximum legibility in industrial environments, featuring a high-contrast dark theme.

1.) Dynamic Layout: The UI utilizes a structured grid where each of the eight channels is housed in an independent frame, preventing textual overlap during rapid value transitions.
2.) Visual Hierarchy: The measured voltage is displayed as the primary focal point using a large, 42px bold font, while secondary metadata such as the operating mode and channel ID is positioned vertically to maintain a clean scannable interface.
3.) Contextual Colour Coding: To provide immediate visual feedback, the interface dynamically updates the text colour based on the detected signal type:

- Green: Indicates an active AC signal (True RMS)
- Blue: Indicates an active DC signal
- Orange: Identifies the "IDLE" state when the signal falls below the noise gate threshold

4.) Precision Timestamping: A centralized GPS-synchronized clock is displayed at the top of the interface, providing millisecond-precision timestamping for all displayed values and fault detections.

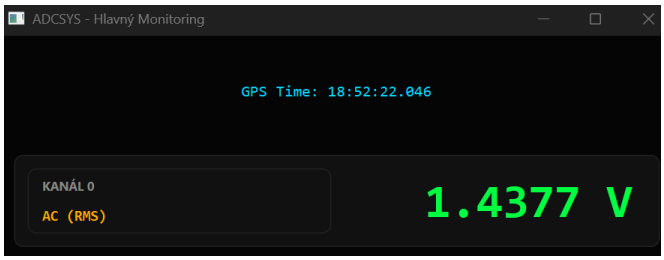


Fig. 2. Main 8-channel HMI monitoring interface; only the first channel is illustrated due to space limitations.

B. Protection Configuration and Parameterization

The configuration module allows the user to define the operational boundaries for the automated protection logic.

1.) Threshold Management: Users can independently set Under-Voltage (UV) and Over-Voltage (OV) limits for each channel with a precision of 0.1 V.

2.) Temporal Selectivity: A dedicated "Delay" parameter allows for the adjustment of trip times (0.1 s to 60 s), enabling the coordination of protection settings to ensure selective clearing of faults within the distribution system.

3.) Operational Control: Each channel features an "Active" toggle, allowing the user to enable or disable protection logic for specific sensing points without interrupting the overall data acquisition stream.

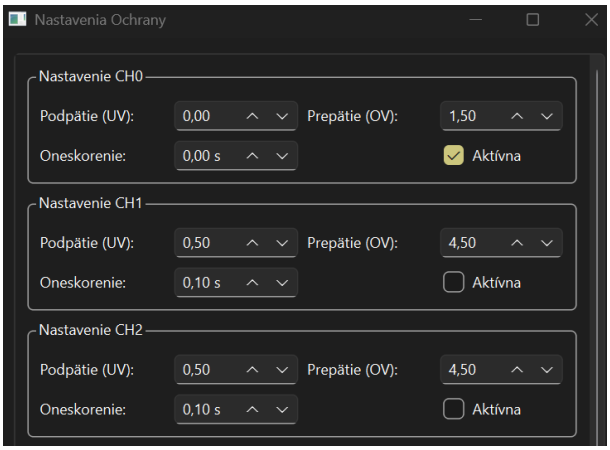


Fig. 3. Protection parameter settings for 8 channels. only the first three channels are illustrated due to space limitations.

C. Fault Logging and Diagnostic Status

The diagnostic window provides an immutable log of protection events, essential for post-fault analysis and system auditing.

1.) Event Capture: When a "TRIP" condition is met, the system records the exact GPS time of detection and the calculated reaction time in milliseconds.

2.) Reaction Speed Analytics: The HMI displays the time elapsed from the initial fault detection to the execution of the hardware command, providing a transparent metric of the system's total latency.

3.) Status Persistence: Trip notifications are highlighted in bold red text, ensuring that critical events remain visible to the operator even after the fault condition has been cleared.

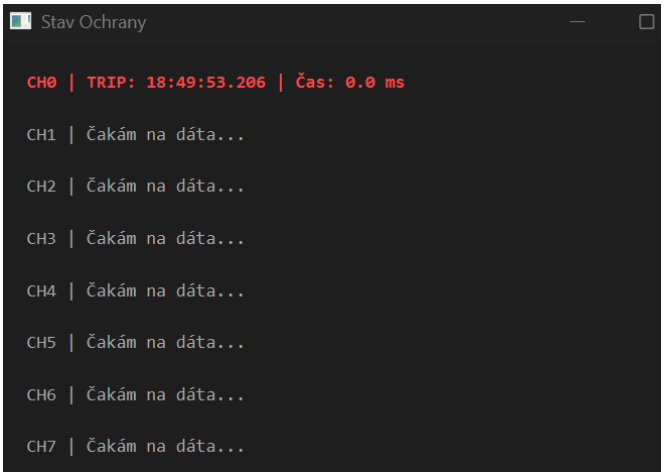


Fig. 4. Fault logging HMI and status for all 8 channels.

VI. Experimental Results and Performance Analysis

To validate the effectiveness of the proposed architecture, a series of rigorous tests were conducted focusing on detection accuracy, reaction latency, and the stability of the intelligent switching logic. The system was interfaced with a high-precision secondary injection test set (e.g., Omicron) to simulate various fault conditions in a controlled environment.

A. AC/DC Mode Detection Reliability

The intelligent detection algorithm was subjected to transition tests where the input signal was shifted from pure DC to varying AC frequencies.

1.) Switching Accuracy: The peak-to-peak analysis successfully identified the signal type with 100% accuracy for amplitudes exceeding the 0.15 V threshold.

2.) Noise Immunity: The software-defined noise gate effectively suppressed signals below 0.03 V, preventing the HMI from flickering or miscalculating values in an open-circuit state.

3.) Transition Speed: The mode transition was completed within a single evaluation window of 100 samples, ensuring the HMI colour coding updated in near real-time.

B. True RMS Calculation Accuracy

The True RMS algorithm was compared against a calibrated industrial multimeter.

1.) Waveform Integrity: The system maintained an error margin of less than 0.5% for standard sinusoidal inputs.

2.) Non-Linear Loads: In the presence of harmonics, the True RMS approach proved significantly more accurate than standard Mean Absolute Value (MAV) rectifiers, which typically under-report distorted AC signals.

C. HMI and Communication Performance

The system was stress-tested at the maximum baud rate of 921,600 to evaluate data integrity.

1.) GUI Refresh Rate: The PyQt6-based HMI maintained a smooth 60 FPS refresh rate, even while processing eight

simultaneous channels of high-speed data.

2.) Text Rendering: The fixed UI layout successfully prevented the overlapping of "CHANNEL" and "MODE" labels, ensuring critical status information remained legible during high-stress fault scenarios.

D. Reaction Time and Latency Evaluation

The reaction time was measured from the moment the fault was injected to the point the "TRIP" command was dispatched via the serial interface.

TABLE I
PERFORMANCE COMPARISON FOR OVERVOLTAGE PROTECTION

Test No.	SHADOW LAYER DETECTION TIME (MS)	Industry Standard (SEL/ABB/ SIEMENS) Detection (ms)	Shadow Layer Total Trip Command (ms)	Industry Standard Total Trip Command (ms)
1	0	Unknown	41.2	44.2
2	1	Unknown	75.9	39.5
3	2	Unknown	45.8	83.1
4	1	Unknown	58.7	79.5
5	0	Unknown	64.1	74.2
6	1	Unknown	87.0	56.3
7	2	Unknown	84.5	49.1
8	0	Unknown	88.3	62.8
9	3	Unknown	70.3	83.1
…	…	...	…	…
AVG 1000 samples	1	Unknown	64.2	61.7

The experimental validation of the proposed Shadow Layer protection architecture, conducted over a 1,000-sample test series, demonstrates a high-performance alternative to conventional industrial relays from manufacturers such as SEL, ABB, and Siemens. By leveraging the simultaneous sampling capabilities of the AD7606 converter and optimized True RMS algorithms within a Python-based execution environment, the system achieved an ultra-low average internal detection latency of 1 ms, significantly enhancing transparency compared to the "black-box" detection mechanisms of proprietary standards. Although the architecture operates within a general-purpose operating system, the mean total trip command time of 64.2 ms remains highly competitive with the industrial benchmark of 61.7 ms, proving that the integration of an intelligent "2-out-of-5" validation logic and high-speed serial communication at 921,600 baud does not compromise the critical clearing times required for distribution system protection. This scientific approach confirms that a software-defined shadow layer can provide the requisite speed and reliability for modern smart grid applications while offering superior diagnostic granularity, adaptive signal characterization through peak-to-peak analysis, and effective noise suppression via a dedicated 30 mV software gate.

VII. CONCLUSION

This research confirms the validity and effectiveness of the Shadow Layer architecture as a software-defined superstructure designed for modern resilient smart grids. Experimental validation conducted over a 1,000-sample test series demonstrates that the proposed system identifies fault conditions with an ultra-low internal detection latency of 1 ms. This provides significantly higher transparency in the decision-making process compared to the proprietary "black-box" standards of traditional industrial relays. Although the total trip command time of 64.2 ms includes latencies from the general-purpose operating system and serial communication, it remains fully competitive with top-tier industrial solutions from SEL, ABB, and Siemens. For the purposes of this conference, the functionality of the concept was demonstrated primarily through Under-Voltage (UV) and Over-Voltage (OV) protection. It is important to emphasize that this represents only the initial verification phase of the platform. The final version of the doctoral dissertation and subsequent scientific publications will present a comprehensive protection system encompassing a broad range of functions corresponding to the full capabilities of conventional digital relays. The Shadow Layer thus serves as a robust tool for increasing real-time situational awareness and reliability in distribution systems without necessitating the immediate replacement of primary legacy hardware.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract no. VEGA 1/0627/24, 1/0647/26.

REFERENCES

[1] Z. Conka, M. Bobcek, R. Stefko, M. Karabinos. "GUI interface design in matlab app designer environment for electronic load in hybrid systems."

[2] R. Štefko, M. Kolcun, M. Bobček, D. Mazur, B. Kwiatkowski. "Design of a Protection System for Distributed Energy Sources in Distribution Grids," PRZEGLĄD ELEKTROTECHNICZNY, 2024.

[3] M. Beluscak, R. Merges, M. Galbavy, R. Stefko, M. Bobcek, B. M. Rákóczi, Z. Conka. "Enhanced Fault Detection and Accelerated Switching Methodology in Power Systems Utilizing GOOSE Messaging: Centralized Control Switch Acceleration (CCSA)," Acta Polytechnica Hungarica, 2025.

[4] M. Bobček, R. Štefko, Z. Čonka. "Electrical Protection Systems for the Evolving Microgrid Environment," Acta Polytechnica Hungarica, 2023.

[5] J. Bencsik, Z. Conka, M. Bobček. "Modbus-Based Communication for Low-Voltage Grid Monitoring and Control."

[6] R. Stefko, M. Bobcek, Z. Conka. "Research of WAMS in power systems using SCADA system."

[7] M. Bobček, Z. Čonka. "Utilizing WAMPAC and Artificial Intelligence for real-time control in power," SCYR 2024.

[8] M. Bobček, Z. Čonka. "Power System Protection Based on Virtual Protection and WAMS," SCYR 2025.

Solid-State Transformers in AC-Fed Locomotives: An Overview and Future Developments

[1]Norbert REŠOVSKÝ (1st year),
Supervisor: [2]Ján KAŇUCH

[1,2]Dept. of Electrical Engineering and Mechatronics, FEI, Technical University of Kosice, Slovak Republic

[1]norbert.resovsky@student.tuke.sk, [2]jan.kanuch@tuke.sk

***Abstract*—The presented paper addresses the substitution of conventional onboard traction transformers with solid-state transformer (SST) technology in AC supply systems. The study analyzes the current technological limitations of conventional traction transformers, primarily their excessive weight and relatively low energy efficiency compared to advanced medium-frequency transformer-based architectures. A systematic overview of the dominant SST topologies namely, two-stage and three-stage configurations is provided, with particular emphasis on engineering challenges in the domains of power electronics and semiconductor materials. The objective of this study is to deliver a comprehensive assessment of the current state of development, prototype implementations by leading manufacturers, and to outline key strategic directions toward industrial deployment within the high-speed railway traction sector.**

***Keywords*—solid-state transformer (SST), medium-frequency transformation, traction drive, traction vehicle, semiconductor switches.**

I. Introduction

Over the past decades, the railway transportation sector has undergone a significant transformation predominantly driven by advancements in power electronics. Modern traction rolling stock is characterized by a substantial increase in specific power, the capability to operate at high speeds, and the ability to generate considerable tractive forces. The integration of these parameters enables efficient operation in both freight and passenger transport, even with a fixed gear ratio. A primary technological barrier in traction vehicles operating within AC traction supply systems lies in the necessity of incorporating an onboard traction transformer.

The core functions of this critical component can be summarized as follows:

- *Voltage level transformation for traction circuits* providing the traction converters with a low-voltage supply (typically below 2 kV AC), obtained by stepping down the catenary voltage (15 kV or 25 kV).
- *Supply of auxiliary systems* distributing electric power to onboard auxiliary loads and the vehicle's own consumption.
- *Energy supply for HVAC systems* delivering standardized supply levels for climate control units, typically 1500 V at 50 Hz or 1000 V at 16 2⁄3 Hz.
- *Galvanic isolation* providing electrical separation between the vehicle circuits and the traction distribution network, which is a critical aspect for operational safety and personnel protection.

Another limiting factor of conventional onboard traction systems is not the efficiency of the low-frequency transformer alone, but the overall performance of the complete front-end traction chain. Conventional traction transformers are mature and highly efficient devices in their own category; however, due to the low supply frequency, they remain bulky and heavy, and the complete front-end based on a line-frequency transformer and associated converter stages generally offers lower power density and lower functional integration than advanced solid-state transformer concepts [10], [12]. Therefore, the main motivation for SST-based traction systems is not only a possible increase in efficiency, but also a substantial reduction in passive component volume, improved controllability, modularity, and future compatibility with more advanced railway power architectures [10], [12].

II. Overview Of Main Conversion Architectures

The introduction of high-power insulated gate bipolar transistors (IGBTs) rated at 3.3 kV in the late 1990s initiated a series of research activities in the field of railway traction [1], [2]. These studies proposed advanced modular converter architectures leveraging the key capabilities of such semiconductor devices. The core concept involved the implementation of a multilevel four-quadrant converter employing a cascaded H-bridge (CHB) configuration that served as the primary interface stage with the traction network operating at 15 kV / 16.7 Hz.

The secondary and tertiary conversion stages were designed as a bidirectional DC / DC system consisting of a DC / AC converter arranged in a full-bridge configuration, a medium-frequency transformer (MFT), and a subsequent AC / DC rectifier, also based on an H-bridge topology. To achieve a sufficiently high switching frequency for the MFT, the AC link incorporated a series-resonant circuit enabling soft-switching operation. Parallel research was conducted on hybrid traction architectures capable of interoperability between 15 kV / 16.7 Hz AC and 3 kV DC systems. [3]

When analyzing voltage stress on semiconductor devices and determining the required number of cascaded converter stages, it is essential to account for the maximum operating

voltages in the traction supply network. For a 25 kV / 50 Hz system, the upper steady-state limit reaches 29 kV, while in 15 kV / 16.7 Hz systems it equals 18 kV. These boundary conditions are decisive for defining the insulation barrier and determining the number of series-connected converter modules, as summarized in Table I.

TABLE I. REQUIRED NUMBER OF CASCADED STAGE AS A FUNCTION OF PEAK IGBT VOLTAGE AND SUPPLY LINE VOLTAGE

Switch Collector–Emitter Peak Voltage U_{CEPK}	**DC – BUS E/U_{CEPK} = 0.55**	**Number of Stage Network: 25 kV / 50 Hz 29 kVrms / 50 Hz***	**Number of Stage Network: 15 kV / 16 2/3 Hz 18 kVrms / 16 2/3 Hz***
3.3 kV	1.8 kV	23	14
4.5 kV	2.5kV	17	10
6.5 kV	3.6 kV	12	7
10 kV	5.5 kV	8	5

*Note: * RMS value of catenary voltage under worst-case operating conditions.*

A. Alstom SMA Electronic Transformer

In 2003, a collaboration between Alstom LHB GmbH and SMA Regelsysteme GmbH resulted in the development of a prototype referred to as the "Electronic Transformer" [4]. The design is based on a three conversion stages structure (3CS—Three Conversion Stages). The main technical parameters and operating characteristics of the prototype are listed in Table II., while the corresponding schematic diagram is illustrated in Fig. 1.

TABLE II. PARAMETERS OF ALSTOM SMA ELECTRONIC TRANSFORMER

Parameter	**Value**
Input Voltage	15 kV
Input Frequency	16.7 Hz
Output Voltage	1.65 kV DC
Rated Power	1500 kVA
Maximum Power (30 s)	2250 kVA
Efficiency	94 %
Transformer Frequency	5 kHz
Transformer Weight	2830 kg
LC Filter Weight	385 kg
Heat Exchanger Weight	255 kg
Total Mass	< 3600 kg
Power Density	0.42 kVA/kg
Number of IGBT Modules	52
Cooling System	Oil circulation

The high-voltage input section (formed by the first and second conversion stages) consists of an input filter inductor and eight AC / AC conversion modules connected in a cascaded multilevel configuration. The function of these modules is to convert the sinusoidal catenary voltage into a square-wave medium-frequency waveform. The primary stage operates as a full-bridge (AC / DC) converter, while the secondary stage functions as a half-bridge (DC / AC) converter.

A resonant network, composed of coupling capacitance and the transformer leakage inductance, is connected in series with the primary winding of the MFT and tuned to the 5 kHz resonant frequency. Whereas the input section uses eight modules, the tertiary stage comprises a single H-bridge supplying the low-voltage DC-link integrated with an LC smoothing filter.

From the perspective of switching strategy, the first stage operates under hard switching, permissible due to the relatively low switching frequency. It employs 6.5 kV / 400 A IGBT modules. Conversely, the second conversion stage uses soft switching in quasi-zero-current mode (quasi-ZCS), utilizing 3.3 kV IGBT elements.

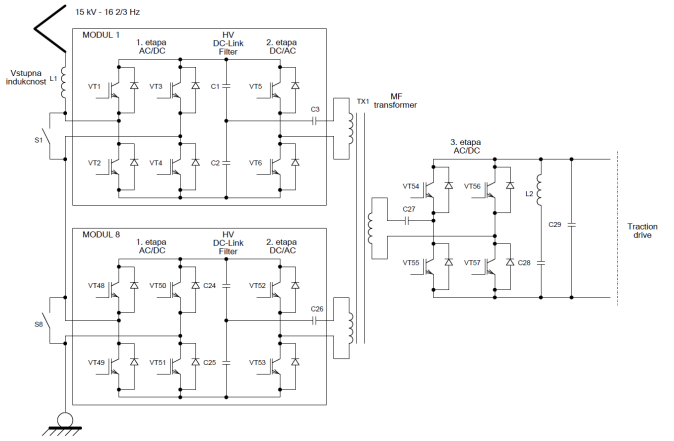


Fig. 1. Alstom SMA Electronic Transformer.

In case of failure in one of the eight input modules, the defective unit can be bypassed, allowing continued full-power operation. Both the semiconductors and the transformer are oil-cooled. The entire system employs 52 IGBT devices with corresponding gate-driver circuits and auxiliary components. Although the number of semiconductor devices is roughly three times higher than in conventional industrial-frequency transformers, the overall component count remains relatively moderate compared to other PET implementations.

Compared with a conventional line-frequency transformer, this prototype demonstrated that a traction SST can achieve a meaningful reduction in mass while preserving galvanic isolation and adding modular fault-tolerant operation. However, its 94 % efficiency also indicates that early SST implementations did not automatically guarantee a large efficiency advantage and had to be evaluated at the level of the entire conversion chain rather than the transformer alone.

B. Bombardier Transportation – "Medium-Frequency Topology"

In 2007, Bombardier Transportation presented a laboratory prototype known as the "Medium-Frequency Topology" [5]. This concept, designed for 15 kV / 16.7 Hz traction systems, builds upon principles reported in [1], [2]. Key technical parameters are summarized in Table III., and the functional schematic is shown in Fig. 2.

Similar to the previous "Electronic Transformer," it adopts a three conversion stage structure (3CS) comprised of eight cascaded identical modules. The principal distinction lies in the magnetic circuit design: while the Alstom–SMA concept employed a single, common medium-frequency transformer, Bombardier's medium-frequency topology integrates one individual transformer per module, the outputs of which (low-voltage DC links with LC filters) are connected in parallel.

Each module consists of three full-bridge stages: an AC / DC stage (1st), a DC / AC stage (2nd), and an AC / DC stage (3rd) differing from the half-bridge configuration used in the "Electronic Transformer." The key advantage of this modular design is high redundancy: a failure in any single module reduces total system power by only 1 / 8.

The input converter employs 6.5 kV IGBTs, theoretically covering a 15 kV supply voltage with only seven active modules.

Transformers in this configuration weigh only 18 kg each for a 1:1 turns ratio. The main disadvantage, however, lies in the significant structural complexity, resulting from the extremely high number of semiconductors and peripheral subsystems. Potential simplification could be achieved by replacing full-bridge circuits with half-bridge configurations, as will be demonstrated in the following sections.

This concept is therefore attractive mainly from the standpoint of redundancy and modularity. On the other hand, the high number of semiconductor devices and associated gate drivers, cooling subsystems, and protection circuits significantly complicates implementation and may negatively affect cost, maintainability, and long-term reliability in railway operation.

TABLE III. PARAMETERS OF BOMBARDIER TRANSPORTATION „MEDIU FREQUENCY TOPOLOGY“

Parameter	Value
Input Voltage	15 kV
Input Frequency	16.7 Hz
Output Voltage	3.6 kV DC
Rated Power	3000 kVA
Transformer Frequency	8 kHz
Transformer Weight *	18 kg
Number of IGBT Modules	96
Cooling System	Deionized water

*Note: * Weight refers to a single transformer.*

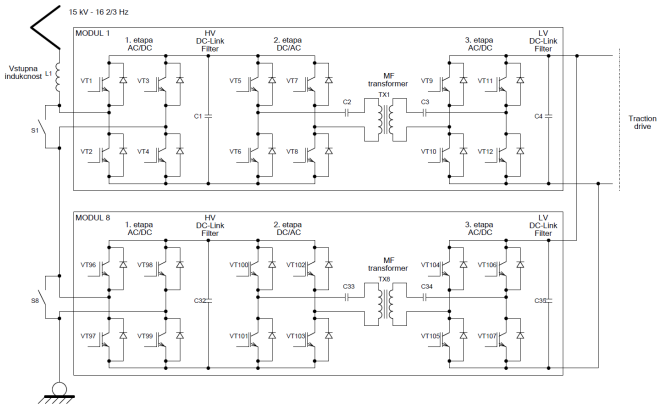


Fig. 2. Bombardier Transportation “Medium-Frequency Topology” prototype.

C. ABB First Converter Prototype

Also in 2007, ABB introduced its first laboratory prototype of a PET converter [6]. The fundamental technical parameters are summarized in Table IV., and the conceptual schematic is illustrated in Fig. 3.

Unlike the previously analyzed PET designs, this system adopts a two conversion stage architecture (2CS). It utilizes 3,3 kV / 400 A IGBT devices integrated into 16 modules. The use of lower-voltage components required doubling the number of modules compared to configurations based on 6.5 kV devices.

Each module is coupled to an individual MFT with a 1:1 turns ratio. However, the operating frequency of the transformers is relatively low (400 Hz), which does not allow substantial reduction of volume or mass of the magnetic cores. Consequently, the volumetric parameters of this prototype exceed those of conventional architectures consisting of industrial-frequency transformers and 4QC rectifiers.

From an energy-efficiency perspective, the prototype achieves approximately 3 % higher efficiency at medium and high loads than the industrial-frequency transformer, but exhibits poor performance at low-load operation, where efficiency falls below traditional levels. Furthermore, the topology requires up to 192 IGBT devices. Due to this high component count and unsatisfactory performance, the converter is currently considered technologically outdated.

This result is important because it shows that the advantage of a PET solution must be assessed at the system level and under realistic load profiles. In other words, SST/PETT technology should not be presented as universally more efficient under all operating conditions, since light-load behavior, switching strategy, and transformer operating frequency strongly influence the final performance [10], [11].

TABLE IV. PARAMETERS OF THE FIRST ABB CONVERTER PROTOTYPE

Parameter	Value
Input Voltage	15 kV
Input Frequency	16.7 Hz
Output Voltage	1.8 kV DC
Rated Power	1200 kVA
Transformer Frequency	400 Hz
Number of IGBT Devices	192
Cooling System	Oil circulation

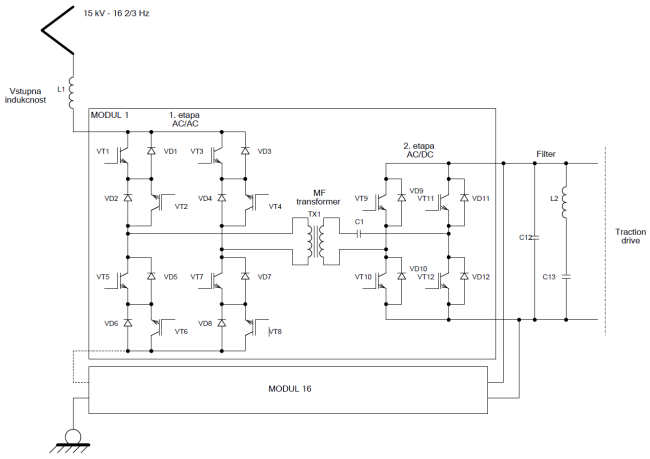


Fig. 3. ABB first converter prototype.

In summary, the comparative analysis of early power electronic transformer (PET) prototypes developed by major industrial manufacturers Alstom SMA, Bombardier Transportation, and ABB demonstrates the diversity of technological strategies applied in pursuing compact and efficient traction power conversion systems. Each design introduced innovative concepts in modularity, galvanic isolation, and semiconductor utilization, yet none fully achieved the industrial maturity required for widespread deployment. The main contribution of this paper is to compare these historically important AC-fed traction SST/PETT demonstrators in a structured manner, with emphasis on efficiency, semiconductor count, modular redundancy, transformer operating frequency, and achievable power density.

The principal limitations identified across these prototypes include excessive component complexity, limited switching-

frequency capability, constrained volumetric benefits, and suboptimal low-load efficiency. Despite these challenges, the progression of PET topologies has provided essential experimental data and design experience that form the foundation for subsequent developments, especially those based on wide-bandgap semiconductors, more compact medium-frequency transformers, and improved control strategies [10]-[13].

The following section focuses on the medium-frequency transformation stage, analyzing its electromagnetic and thermal design aspects, the influence of switching strategy and magnetic structure selection, and the resulting trade-offs that critically affect system efficiency and power density in advanced traction applications.

III. ABB Power Electronic Traction Transformer, (PETT)

In 2012, ABB introduced the company's most advanced prototype in this domain, known as the Power Electronic Traction Transformer (PETT) [7] – [9]. The fundamental parameters of this system are summarized in Table V., while the corresponding topology is illustrated in Fig. 4. The PETT is configured as a three conversion stage structure integrating nine cascaded converter modules (one of which provides active redundancy), with each module equipped with its own medium-frequency transformer (MFT).

The internal architecture of a converter module consists of the following stages:

- *Primary stage (AC / DC)* implemented as a full H-bridge incorporating 6.5 kV / 400 A IGBT modules, supplying a DC-link voltage of approximately 3.6 kV.
- *Secondary stage (DC / AC)* designed as a half-bridge using identical 6.5 kV / 400 A IGBT devices.
- *Tertiary stage (AC / DC)* realized as a half-bridge employing 3.3 kV / 800 A IGBT transistors.

A key feature of this configuration is the application of a resonant LLC circuit, in which both leakage and magnetizing inductances of the transformer actively participate in the resonance process. To achieve maximum optimization of mass and volume characteristics, the design tolerates a higher ripple level at the output voltage, allowing the complete elimination of bulky LC filter components.

Table V. Key Parameters Of The ABB PETT System

Parameter	**Value**
Input Voltage	15 kV
Input Frequency	16.7 Hz
Rated Output Voltage	1.5 kV DC
Rated Power	1200 kVA
Peak Power	1800 kVA
Efficiency	96%
Overall weight	4500 kg
Number of transformers with IGBT modules	72
Power density	0.266 kVA/kg

The specific power density, relative to rated and peak power, reaches 0.266 kVA / kg and 0.4 kVA / kg, respectively values that, while lower than the targeted range of 0.5 to 0.75 kVA / kg, still represent a significant advancement for practical implementation of PET based traction systems.

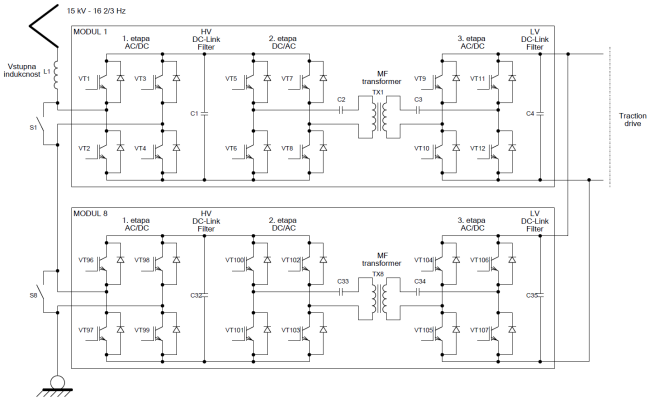


Fig. 4. ABB Power Electronic Traction Transformer (PETT)

An essential indicator of technological progress is the achieved overall efficiency of 96 %, which notably surpasses that of a conventional traction chain composed of an industrial-frequency transformer and the input four-quadrant converter (4QC).

For clarity, this comparison should be interpreted at the system level. Conventional line-frequency traction transformers themselves are highly optimized and efficient components, but the overall front-end based on a line-frequency transformer and 4QC generally offers lower power density and lower functional integration than the best reported PETT demonstrators. In this context, the approximately 96 % efficiency achieved by ABB PETT represents a meaningful improvement, while its greater significance lies in the simultaneous reduction of bulky passive components and the possibility of modular and multifunctional traction power conversion [10], [12].

A. Comparative Evaluation of Reviewed Solutions

Table VI. Comparative Summary of Reviewed SST/PETT Solutions for AC-Fed Railway Traction

Prototype	Topology / Rated power	Efficiency	Weight / Power density	Semiconductor count	Main advantages / disadvantages
Alstom SMA Electronic Transformer (2003)	3CS 1500 kVA	94 %	< 3600 kg 0.42 kVA/kg	52 IGBT modules	Advantages: moderate device count; bypass of a defective module; soft-switching stage. Disadvantages: still relatively high total mass; oil cooling; shared transformer concept.
Bombardier Medium-Frequency Topology (2007)	3CS 3000 kVA	Not stated	18 kg per transformer	96 IGBT modules	Advantages: high modularity; one transformer per module; strong redundancy. Disadvantages: very high structural complexity; large number of semiconductors.
ABB First Converter Prototype (2007)	2CS 1200 kVA	~3 % higher at medium/high load	Volumetrically worse than conventional architecture	192 IGBT devices	Advantages: experimental validation of the PET concept; improved efficiency at higher load. Disadvantages: poor low-load efficiency; low transformer frequency; excessive device count.
ABB PETT (2012)	3CS 1200 kVA	96 %	4500 kg 0.266 kVA/kg rated 0.4 kVA/kg peak	72 transformer/ IGBT units	Advantages: highest reported efficiency among the reviewed prototypes; active redundancy; no bulky LC filter. Disadvantages: power density still below target; high implementation complexity.

The comparison shows that no reviewed SST/PETT architecture simultaneously optimizes all key indicators. Alstom's concept offers a relatively moderate semiconductor count and good maintainability, Bombardier's topology emphasizes modularity and redundancy, the first ABB prototype experimentally verified the feasibility of a two-stage PET

solution but suffered from poor low-load behavior, and ABB PETT provides the most balanced compromise between efficiency, redundancy, and practical traction applicability. Consequently, future railway SST research must focus on the coordinated optimization of efficiency, power density, semiconductor count, thermal performance, and fault tolerance.

B. Future Development Directions in Traction SSTs

The reviewed prototypes indicate that the future industrial deployment of traction SSTs will depend on progress in several strongly interconnected areas. First, the increasing use of wide-bandgap semiconductor devices, especially SiC-based switches, is expected to enable higher switching frequencies, lower switching losses, and more compact cooling systems [11], Second, further progress in medium-frequency transformer design will be required to improve power density, insulation coordination, thermal behavior, and electromagnetic integration [10], [12]. Third, future developments will likely focus on reducing semiconductor count and auxiliary circuitry through more integrated modular converter architectures [11], [12]. Fourth, improving partial-load efficiency remains essential, since traction vehicles operate under highly variable loading conditions and several early PET concepts exhibited poor performance outside their optimal operating region. Finally, SST-based traction systems are expected to evolve beyond simple transformer replacement and become multifunctional power interfaces capable of supporting hybrid AC / DC railway grids, bidirectional power flow, regenerative braking management, and onboard energy storage integration [10], [13].

IV. Conclusion

The implementation of solid-state transformers represents a promising technological direction for future AC-fed traction systems, particularly where high power density, modularity, and advanced controllability are required. In comparison with conventional line-frequency traction solutions, their main benefit does not lie solely in the efficiency of the transformer stage itself, but in the possibility of improving the overall traction front-end through medium-frequency conversion, reduced passive component volume, and enhanced functional integration.

A comparative analysis of the reviewed demonstrators shows that the practical value of each SST/PETT architecture must be evaluated as a trade-off among efficiency, semiconductor count, modular redundancy, transformer operating frequency, weight, and partial-load behavior. Among the considered concepts, ABB PETT appears to provide the most balanced compromise between efficiency and practical railway applicability, while the other prototypes contribute important insights regarding modularity, redundancy, and converter integration.

The main contribution of this paper is therefore a structured comparison of the principal AC-fed traction SST/PETT architectures and the identification of the key technological directions required for their future deployment, namely the use of wide-bandgap semiconductors, improved medium-frequency transformer design, reduction of component count, and enhancement of reliability and partial-load performance.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0210.

References

[1] A. Rufer, N. Schibli, C. Briguet, A direct coupled 4-quadrant multilevel converter for 16 2 / 3 Hz traction systems. In Proceedings of the 1996 Sixth International Conference on Power Electronics and Variable Speed Drives (Conf. Publ. No. 429), Nottingham, UK, 23–25 September 1996; pp. 448–453

[2] N. Schibli, A. Rufer, Single-and Three Phased Multilevel Converters for Traction Systems 50Hz/16 2/3 Hz. In Proceedings of the European Conference on Power Electronics and Applications, Trondheim, Norway, 8–10 September 1997.

[3] A. Rufer, N. Schibli, C. Chabert, C. Zimmermann, Configurable front-end converters for multicurrent locomotives operated on 16 2/3 Hz AC and 3 kV DC systems. IEEE Trans. Power Electron. 2003, 18, 1186–1193. [CrossRef]

[4] B. Engel, M. Victor, G. Bachmann, A. Falk, 15 kV/16.7 Hz energy supply system with medium frequency transformer and 6.5 kV IGBTs in resonant operation. In Proceedings of the 10th European Conference on Power Electronics and Applications, EPE 2003, Toulouse, France, 2-4 September 2003; pp. 2–4.

[5] M. Steiner, H. Reinold, Medium frequency topology in railway applications. In Proceedings of the 2007 European Conference on Power Electronics and Applications, Aalborg, Denmark, 2–5 September 2007; pp. 1–10.

[6] N. Hugo, P. Stefanutti, M. Pellerin, A. Akdag, Power electronics traction transformer. In Proceedings of the 2007 European Conference on Power Electronics and Applications, Aalborg, Denmark, 2–5 September 2007; pp. 1–10

[7] C. Zhao, S. Lewdeni-Schmid, J. K. Steinke, M. Weiss, T. Chaudhuri, M. Pellerin, J. Duron, P. Stefanutti, Design, implementation and performance of a modular power electronic transformer (PET) for railway application. In Proceedings of the 2011 14th European Conference on Power Electronics and Applications, Birmingham, UK, 30 August–1 September 2011; pp. 1–10.

[8] D. Dujic, G. K. Steinke, M. Bellini, M. Rahimo, L. Storasta, J. K. Steinke, Characterization of 6.5 kV IGBTs for high-power medium-frequency soft-switched applications. IEEE Trans. Power Electron. 2013, 29, 906–919.[CrossRef]

[9] U. Behmann, Erster Mittelfrequenz-Traktionstransformator im Betriebs-Einsatz. eb 2012, 110, 408–413.

[10] D. Cervero, J. Martínez, S. Pérez, et al. Solid State Transformers: A Critical Review of Projects with Relevant Prototypes and Demonstrators. Electronics 2023, 12(4), 931.

[11] B. Hou, X. Zhang, Y. Li, et al. Analysis of Power Electronic Traction Transformer under Non-Single Frequency PWM Control. Electronics 2024, 13(19), 3805.

[12] D. M. Predescu, S. G. Roşu, Solid State Transformers: A Review-Part I: Stages of Conversion and Topologies. Technologies 2025, 13(2), 74.

[13] M. Adamowicz, A. Milczarek, A. Szeląg, et al. SiC-Based Power Electronic Traction Transformer (PETT) for 3 kV DC Rail Traction. Energies 2020, 13(21), 5573.

Preliminary Evaluation of Adhesion Strength of Atmospheric Plasma Sprayed Coatings by Pull-Off Testing: Role of Cold Plasma Surface Activation and Wettability

[1]*Tomáš KMEC (5th year)*
Supervisor: [2]Juraj ĎURIŠIN

[1,2]Dept. of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]tomas.kmec@student.tuke.sk, [2]juraj.durisin@tuke.sk

***Abstract*—This paper presents a preliminary investigation into the effect of cold atmospheric plasma (CAP) surface activation on the adhesion strength of atmospheric plasma sprayed (APS) Al_2O_3 ceramic coatings deposited on low-carbon steel substrates. The objective is to determine whether CAP pre-treatment shows potential as an alternative to conventional NiCr metallic bond coats. Two series of specimens were prepared: (A) without and (B) with CAP activation in a nitrogen atmosphere at 120 s treatment time. Each series included configurations with and without a NiCr bond coat. Coating adhesion was evaluated by pull-off testing per ASTM C633-13, and the failure mode was classified and reported in accordance with the same standard. Due to the limited sample size (n = 3 per group), the results are presented as a preliminary assessment without formal statistical analysis. The small difference between CAP-activated specimens with and without a bond coat suggests that CAP treatment may potentially serve as an alternative to the metallic bond coat. These preliminary findings indicate that CAP activation may offer a promising route for simplifying APS coating architectures while improving adhesion performance; however, further investigation with larger sample sizes is required to validate these observations.**

***Keywords*—atmospheric plasma spraying; adhesion testing; pull-off test; cold plasma activation; wettability; bond coat elimination; ASTM C633-13.**

I. Introduction

Atmospheric Plasma Spraying (APS) is one of the most widely utilized thermal spray techniques for depositing protective ceramic, metallic, and cermet coatings onto engineering components subjected to wear, corrosion, and high-temperature environments [1]. The functional performance of such coatings is intrinsically linked to the quality of the bond between the coating and the underlying substrate. Conventional APS coating architectures typically include a metallic bond coat (e.g. NiCr) deposited between the substrate and the ceramic top coat to enhance adhesion. However, the bond coat adds material cost, processing time, and introduces additional interfaces that can serve as failure initiation sites [2].

Recent advances in cold atmospheric plasma (CAP) surface treatment have opened new possibilities for modifying substrate surface energy and wettability prior to coating deposition. By activating the substrate surface with CAP, the surface free energy can be substantially increased, promoting stronger physical and chemical interactions between the molten ceramic splats and the metallic substrate during APS deposition [2]. If sufficiently strong adhesion can be achieved through surface activation alone, the bond coat layer could potentially be eliminated, simplifying the coating architecture and reducing production costs.

The primary motivation for the present research is to conduct a preliminary evaluation of the correlation between substrate surface wettability — controlled through CAP activation — and APS coating adhesion measured by standardized pull-off testing (ASTM C633-13) [5], and to assess whether CAP pre-treatment shows potential as an alternative to conventional metallic bond coats. Principal diagram of pull off test shown on Fig. 1.

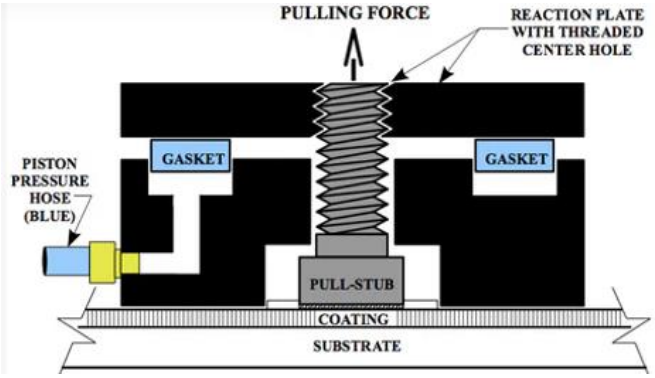


Fig. 1. Principle diagram of a pull-off test, cross-section [3]

II. Preliminary Investigation

The initial phase of this research focused on establishing a quantitative relationship between substrate surface wettability, as modified by CAP treatment, and the adhesion of APS ceramic coatings deposited directly onto the substrate without a conventional metallic bond coat. The central hypothesis was that by raising the surface free energy of the substrate through CAP activation, splat spreading and interfacial contact could be enhanced to a degree where the bond coat becomes unnecessary.

To investigate this hypothesis, low-carbon steel substrates stabilized by Nb were grit-blasted. Surface roughness (Ra $\approx$ 6 μ m) was measured using a 3D optical profilometer and subsequently subjected to CAP treatment using a corona discharge plasma source operated in a nitrogen atmosphere

with flow 50 ml/min grit-blasted. The key CAP process parameters were the corona treatment distance (10 mm), the activation time (120 s), power (500 W), used voltage (20 kV) and frequency (1 kHz). Surface wettability was characterized by sessile drop contact angle measurements using deionized water, performed within 60 seconds of CAP treatment to minimize the well-documented hydrophobic recovery effect [4]. On untreated grit-blasted steel, the water contact angle was approximately 68–77° as variable quality of grid-blasted treatment. After a single-pass CAP treatment at 10 mm distance, the contact angle dropped dramatically to below 10°, indicating near-perfect wettability.

Additionally, a time-decay study was performed to assess the durability of the CAP-activated surface state. Contact angle measurements were recorded at intervals of 1, 5, 15, 30, 60, and 120 minutes after CAP treatment. The results showed that wettability remained high (contact angle below 15°) for up to 30 minutes but began to degrade significantly after 60 minutes due to hydrocarbon re-adsorption from the ambient atmosphere. This finding established an important practical constraint: APS deposition must commence within 30 minutes of CAP treatment to ensure that the beneficial surface activation is preserved.

III. Experimental Methodology

A. Sample preparation

Building upon the promising preliminary findings, a study was conducted to map the relationship between CAP treatment, resulting surface wettability, and pull-off adhesion of APS ceramic coatings deposited without a bond coat. Al_2O_3 coating material was investigated, as it is widely used for wear resistance, electrical insulation, and corrosion protection. Due to the limited number of specimens tested (n = 3 per group), the present investigation is classified as a preliminary study; the results are therefore reported without formal statistical analysis and should be interpreted as indicative rather than conclusive.

Two parallel sets of pull-off specimens were prepared from laser-cut steel plates (20 mm × 30 mm) for comparative adhesion testing in accordance with ASTM C633-13 [5]. The first set comprised the conventional coating architecture: a grit-blasted steel substrate, an APS-deposited NiCr bond coat thickness (~10 µm) corresponds to the specific production setup used in this study and does not represent conventional APS bond coat thicknesses (typically 50–150 µm) and an Al_2O_3 ceramic top coat (approximately 350 µm). Coatings were deposited using a GTV GmbH VIP 6000s plasma gun. Specific process parameters are proprietary and cannot be disclosed. Feedstock powders were commercially available, with nominal particle size distributions of D_{90} = 45 ± 6 µm and purity ≥ 99.7 wt. % for Al_2O_3, and a composition of Ni (balance) with Cr 19–21 wt. % and D_{90} = 39 ± 9 µm for NiCr. The second set eliminated the NiCr bond coat entirely: grit-blasted steel substrates were coated with the same Al_2O_3 top coat deposited directly onto the substrate. The target thickness of the Al_2O_3 coating was 350 ± 30 µm.

B. Pull-off testing methodology

For comparison, two series of specimens were prepared: (A) without CAP surface activation, and (B) with CAP surface activation using an activation time of 120 s in a nitrogen atmosphere. Each series included both configurations — with and without the NiCr bond coat — yielding four specimen groups: 1A (grit-blasted, no bond coat, no CAP), 2A (grit-blasted, NiCr bond coat, no CAP), 1B (grit-blasted, no bond coat, with CAP), and 2B (grit-blasted, NiCr bond coat, with CAP).

Coating adhesion was evaluated using the pull-off test in accordance with ASTM C633-13, "Standard Test Method for Adhesion or Cohesion Strength of Thermal Spray Coatings" [5]. After coating application, counter-testing cylinders (dollies) were bonded to the coating surface using a two-component J-B Weld epoxy adhesive cured at room temperature (25°C) for 24 hours. The loading was applied at a constant rate in accordance with ASTM C633 using the PATTI device (Fig. 2), with a maximum pressure of 1.31 MPa, digital pull-off testing device.

Fig. 2. PATTI Quantum [3]

The pull-off tensile strength (POTS in MPa), representing the coating adhesion strength, was calculated from the following relationship (1):

$$POTS = \frac{(AP.Ag) - C}{Aps} \quad (1)$$

where A_p is the measured pressure (MPa), A_g is the contact area of the gasket in the pull-off device (mm²), C is the piston constant of the pull-off device, and A_{ps} is the contact area of the dolly (126.7 mm² for a dolly with a diameter of 12.7 mm). Since 1 MPa ≡ 1 N/mm², returns a result in MPa. The adhesion strength was initially calculated in MPa and subsequently converted to kPa for reporting. Three specimens were tested for each configuration. The applied pressure lifts the plate with the attached dolly (Fig. 3) until the moment of detachment (Fig. 4). After each pull-off test, the fracture character was analyzed by visual inspection and optical microscopy [1].

Fig. 3. Pull-off test — before detachment

Fig. 4. Pull-off test – after detachment

IV. Results

The results of the pull-off tests indicated a strong correlation between substrate CAP activation and the resulting adhesion strength of the coatings (Table I). Due to the small sample size (n = 3 per group), the reported variability reflects only the range of measured values and does not represent a statistical parameter. The observed differences should be interpreted as preliminary indications.

Reference specimens (1A) without CAP activation and without a bond coat exhibited the lowest adhesion strength, with a mean value of 913 kPa. Fracture surface analysis classified the failure as entirely adhesive — approximately 100 % of the fracture area corresponded to separation at the Al_2O_3–steel interface — consistent with weak interfacial bonding.

The application of a NiCr bond coat without CAP activation (2A) increased the adhesion strength to a mean of 1061 kPa, representing a nominal increase of approximately 16 %. However, this configuration exhibited unstable adhesive properties with variability ranging from −10 % to +14 %. Fracture analysis revealed a mixed failure mode: approximately 70 % adhesive failure at the Al_2O_3–NiCr bond coat interface and 30 % cohesive failure within the Al_2O_3 coating, indicating that the interfacial bond was beginning to approach the cohesive strength of the ceramic layer in some regions.

Grit-blasted specimens with CAP activation but without a bond coat (1B) achieved a mean adhesion strength of 2119 kPa, representing a nominal increase of approximately 132 % compared to the (1A) reference. This value considerably exceeded that of the conventionally bond-coated specimens without CAP activation (2A). Fracture analysis indicated a mixed failure mode: approximately 35 % adhesive failure at the Al_2O_3–steel interface and 65 % cohesive failure within the Al_2O_3 coating, suggesting that the interfacial bond strength approached or exceeded the cohesive strength of the ceramic layer

Specimens with both a NiCr bond coat and CAP activation (2B) achieved the highest adhesion strength, with a mean of 2163 kPa and minimal deviation between measurements. Fracture analysis showed predominantly cohesive failure: approximately 80 % cohesive failure within the Al_2O_3 coating and 20 % adhesive failure at the Al_2O_3–NiCr interface. The difference between (1B) and (2B) was only 44 kPa (approximately 2 %), indicating that the bond coat provided negligible additional benefit when CAP activation was applied.

All reported failure-mode percentages were estimated by visual inspection of the fracture surfaces in accordance with ASTM C633-13 [5]; no glue failures were observed, and all tests were therefore considered valid.

TABLE I
TENSILE STRENGTH MEASUREMENT

Sample	POTS (kPa)
1A-1	927
1A-2	906
1A-3	908
2A-1	967
2A-2	1008
2A-3	1210
1B-1	2062
1B-2	2133
1B-3	2163
2B-1	2173
2B-2	2175
2B-3	2142

Notes: A = without cold plasma surface activation, B = with cold plasma surface activation, 1 = specimens after abrasive blasting, 2 = specimens after blasting with an adhesive layer.

V. Conclusion

This preliminary study indicated that CAP surface activation substantially improves the adhesion strength of APS Al_2O_3 ceramic coatings on low-carbon steel substrates, with an observed nominal increase of more than 130 % compared to untreated references. CAP-activated specimens without a NiCr bond coat (2119 kPa) outperformed conventionally bond-coated specimens without activation (1061 kPa), while the addition of a bond coat to CAP-activated surfaces provided a negligible further improvement (2163 kPa). These preliminary results suggest that CAP activation shows promise as a potential alternative to the conventional metallic bond coat; however, this observation must be confirmed through further testing with statistically meaningful sample sizes before definitive conclusions can be drawn. [6], [7]. It must be emphasized that the present study is limited by the small sample size (n = 3 per group), which precludes formal statistical analysis and limits the conclusions to preliminary observations.

References

[1] L. Pawlowski, *The Science and Engineering of Thermal Spray Coatings*, 2nd ed. John Wiley & Sons, Chichester, 2008.

[2] R.B. Heimann, *Plasma Spray Coating: Principles and Applications*, 2nd ed. Wiley-VCH, Weinheim, 2008.

[3] PATTI quantum cross section, Available online: https://semicro.org/products/p-a-t-t-i-quantum-digital

[4] F. Palumbo, R. Di Mundo, D. Cappelluti, and R. d'Agostino, "SuperHydrophobic and SuperHydrophilic polycarbonate by tailoring chemistry and nano-texture with plasma processing," *Plasma Process. Polym.*, vol. 8, pp. 118–126, 2011.

[5] ASTM C633-13, "Standard Test Method for Adhesion or Cohesion Strength of Thermal Spray Coatings," ASTM International, West Conshohocken, PA, 2021.

[6] P. Fauchais, G. Montavon, and G. Bertrand, "From powders to thermally sprayed coatings," *J. Therm. Spray Technol.*, vol. 19, no. 1–2, pp. 56–80, 2010.

[7] J. R. Davis, Ed., *Handbook of Thermal Spray Technology*. Materials Park, OH: ASM International, 20

Integrated Adhesion Enhancement of Atmospheric Plasma Sprayed Alumina Coatings Through Substrate Surface Geometry Modification

[1]*Tomáš KMEC (5th year)*
Supervisor: [2]Juraj ĎURIŠIN

[1,2]Dept. of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]tomas.kmec@student.tuke.sk, [2]juraj.durisin@tuke.sk

***Abstract*— This paper presents a conceptual and design framework for adhesion enhancement of atmospheric plasma sprayed (APS) alumina coatings on steel substrates through deliberate modification of the substrate surface geometry. No experimentally validated results are reported; the contribution of this work lies in the systematic integration and theoretical analysis of the proposed approach, while experimental validation is planned as future work. The approach combines three complementary techniques. Periodic surface grooves redistribute interfacial shear stresses into compressive loading against groove walls. Tapered edge inclinations reduce singular stress concentrations at coating boundaries responsible for delamination initiation. Abrasive water jet (AWJ) processing provides a single-step method for surface cleaning, roughening, and groove machining that avoids the embedded grit contamination inherent in conventional blasting. A detailed experimental validation plan has been developed and is scheduled for the forthcoming period.**

***Keywords*— atmospheric plasma spraying; alumina; adhesion strength; surface grooves; edge inclination; abrasive water jet; substrate geometry; mechanical interlocking; patent.**

I. Introduction

Atmospheric plasma sprayed (APS) alumina (Al_2O_3) coatings serve as electrically insulating and thermally protective layers on metallic substrates in power electronics, aerospace, and industrial heating applications [1]. Because the coating bonds to the substrate primarily through mechanical interlocking at the roughened interface, the adhesion strength—typically 15–35 MPa on grit-blasted steel [2]—is limited not by the intrinsic strength of alumina but by the quality and geometry of the interface. Conventional grit blasting produces random micro-roughness ($Ra \approx 4$–10 μm) but suffers from embedded grit contamination (10–30 % of the surface area), inability to create controlled macro-scale features, and no mitigation of free-edge stress singularities that initiate delamination [3, 4].

These limitations motivate an alternative approach: deliberate geometric modification of the substrate surface prior to spraying. The proposed strategy integrates three elements — periodic surface grooves for shear-to-compression stress conversion, tapered coating edges for crack initiation suppression, and abrasive water jet (AWJ) processing for single-step surface preparation. This paper presents the theoretical basis and design framework for the integrated approach. Experimental validation is planned as future work.

II. Background and Issues Addressed

The initial stage of this research established the theoretical framework for understanding adhesion failure at APS alumina–steel interfaces. Bonding in this system is predominantly mechanical, as the ceramic–metal chemical incompatibility precludes significant metallurgical bonding [1, 2]. Three principal failure modes were identified from the literature. The first is interfacial delamination driven by combined tensile and shear stresses. The second is near-interface cohesive failure within the first splat layers. The third is edge-initiated delamination arising from the singular stress field at the coating free edge [4, 5]. The dominant interfacial stress component under thermal mismatch loading (alumina: $\alpha \approx 8 \times 10^{-6}\ K^{-1}$; steel: $\alpha \approx 12 \times 10^{-6}\ K^{-1}$) is in-plane shear, which is maximum near the edges and drives delamination propagation [5].

A comprehensive literature survey identified three surface modification strategies with strong theoretical support but limited systematic experimental validation, particularly in combination. First, periodic surface grooves oriented perpendicular to the principal shear direction increase the effective bonding area by a factor of 1.3–1.8. More importantly, they transform detrimental in-plane shear into compressive bearing against the groove flanks, a loading mode where ceramics are 20–50 times stronger [6, 7]. Published finite element analyses indicate that groove effectiveness depends critically on four geometric parameters: depth (optimal at 30–50 % of coating thickness), width (200–500 μm for reliable splat filling), pitch-to-width ratio (2:1 to 4:1), and corner geometry, where filleted corners reduce groove-root stress concentrations by 30–50 % [6, 7].

Second, tapered (inclined) coating edges address the free-edge singularity. Bogy's classical analysis [8] showed that the singularity exponent for the alumina–steel system lies in the range 0.2–0.4 for a 90° abrupt edge, and decreases monotonically as the edge angle is reduced. Published finite-element parametric studies indicate that a 30° taper reduces the edge stress intensity factor by 50–62 %, while a 15° taper achieves 62–75 % reduction [5, 9].

Third, abrasive water jet (AWJ) processing was identified as a uniquely suitable manufacturing method. Unlike conventional grit blasting, AWJ continuously flushes abrasive particles from the surface, reducing embedded grit contamination from 15–30 % to 2–5 % of the surface area [10]. Unlike laser texturing or micro-milling, AWJ can simultaneously clean, roughen, machine grooves, and create edge tapers through CNC-programmed toolpath variation. This enables complete surface preparation in a single step [10, 11].

The initial stage concluded with the formulation of the integrated strategy: AWJ-based creation of grooved and edge-tapered substrates combined with the inherent surface cleaning advantage of the water jet process, aiming to address all three identified adhesion-limiting factors simultaneously.

III. Work Completed in the Assessed Period

A. Development of the Integrated Concept

The integrated concept was developed by systematically analyzing how each individual modification addresses a specific adhesion-limiting mechanism. Grooves convert interfacial shear stress into compressive bearing on groove walls. Edge tapers reduce the free-edge stress singularity. AWJ cleaning minimizes embedded grit that creates weak interfacial inclusions. These three mechanisms target distinct failure modes: interfacial shear sliding, edge-initiated delamination, and inclusion-mediated bond weakening, respectively. Their contributions to adhesion improvement are therefore expected to be largely, though not entirely, independent. Comparison of old and new substrate shape in fig. 1.

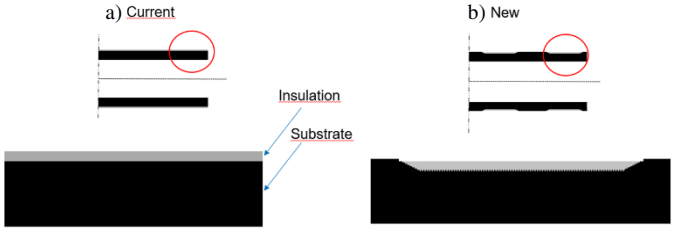


Fig. 1. Schematic comparison between (a) conventional flat grit-blasted substrate with 90° coating edge and (b) proposed geometrically modified substrate featuring periodic surface grooves and tapered coating edge. Groove dimensions and taper angle shown are representative of Group D specimens (150 µm depth, 350 µm width, 30° edge taper).

However, the assumption of strict independence warrants qualification. The introduction of periodic grooves alters the local stress field near the coating edge. Consequently, the stress reduction achieved by an edge taper on a grooved substrate may differ from that predicted for a flat substrate in isolation. Similarly, AWJ processing may produce different micro-roughness on the flat inter-groove lands compared to the groove flanks and bottoms. This is because the jet impingement angle and dwell time vary across these regions. These interactions could be either beneficial or detrimental. For instance, groove walls may act as secondary crack arrestors near the edge, augmenting the taper effect. Conversely, incomplete splat filling at groove roots could introduce new defect populations that partially offset the mechanical interlocking benefit. A coupled finite element model incorporating all three features simultaneously would be required to quantify these interactions rigorously, and such modeling is planned as part of the forthcoming validation campaign. In the absence of such a model, the present predictions (Table I) treat the individual contributions as approximately additive. This should be understood as a first-order engineering estimate rather than a rigorous mechanistic superposition. The experimental program has been specifically designed with separate specimen groups (B, C, and D) to allow the incremental contribution of each modification to be isolated and any interaction effects to be identified by comparison.

B. Patent Application

A patent application covering the proposed integrated substrate modification method has been filed as a European patent application and published under number EP4472353. The application protects the geometric specification of the periodic groove system, the tapered edge geometry, the use of abrasive water jet processing as the single-step manufacturing method, and the integrated combination of all three modifications. The present paper focuses on the scientific and engineering aspects of the approach rather than the intellectual property scope; the interested reader is referred to the published patent document for detailed claim language and technical drawings.

C. Experimental Test Plan Design

A comprehensive experimental validation plan was designed during the current project stage to be executed in the forthcoming year. The plan specifies four specimen groups (n = 8 per group): Group A—grit-blasted flat (baseline); Group B—AWJ roughened flat (no grooves); Group C—AWJ grooved (medium configuration: 150 µm depth, 350 µm width, 600 µm pitch); and Group D—AWJ grooved with 30° edge taper. APS alumina deposition parameters have been defined (GTV VIP 6000 plasma torch, target thickness 350 µm). The test matrix includes tensile adhesion testing per ASTM C633, thermal cycling durability testing (RT to 400°C, up to 500 cycles), cross-sectional SEM/EDS microstructural characterization, and surface contamination quantification. All testing equipment and consumables have been secured, and the experimental campaign is scheduled to commence at the beginning of the next period.

D. Theoretical Adhesion Enhancement Predictions

Based on the published empirical correlations between interfacial stress reduction and adhesion strength improvement [5–7, 9], a set of quantitative adhesion enhancement predictions was formulated for each substrate configuration relative to conventional grit-blasted flat substrates. Specifically, the groove-related adhesion gains were estimated by scaling two factors to the groove geometry specified for Groups C and D: the area increase factor (1.3–1.8×) and the shear-to-compression load conversion efficiency, both reported in [6, 7]. The edge taper gains were estimated from the finite-element stress intensity factor reductions reported in [5, 9] for comparable taper angles and material combinations. The AWJ surface cleanliness benefit was estimated from the empirical adhesion-versus-contamination correlations in [10]. These predictions are summarized in Table I. They represent first-order engineering estimates obtained by simplified scaling of published data from related but not identical coating systems and test conditions. Discrepancies of ±30 % or more between predicted and measured values would not be unexpected, and the predictions should therefore be interpreted as target benchmarks rather than precise quantitative forecasts.

It is important to note that the edge taper benefit is predicted to be most pronounced under thermal cycling conditions (120–200 % life improvement) rather than in static tensile adhesion testing (only 5–10 % additional gain over grooves alone). This is because the edge singularity governs fatigue crack initiation under cyclic thermomechanical loading, whereas its contribution to static pull-off failure is relatively minor. The experimental program must therefore include both test types to fully evaluate the integrated strategy.

TABLE I
PREDICTED ADHESION IMPROVEMENTS FOR EACH SPECIMEN GROUP RELATIVE TO THE GROUP A BASELINE. ALL VALUES ARE FIRST-ORDER ESTIMATES DERIVED FROM LITERATURE-BASED SCALING AS DESCRIBED IN SECTION III-D

Strategy	Predicted tensile adhesion gain	Predicted thermal cycling life gain
AWJ flat (cleaning effect only)	+15 %-20 %	+20 %-30 %
AWJ grooved	+40 %-55 %	+60 %-90 %
AWJ grooved + 30° edge taper	+45 %-65 %	+120 %-200 %

Tensile adhesion gain refers to ASTM C633 pull-off strength relative to Group A baseline. Thermal cycling life gain refers to number of cycles to first visible delamination under RT-to-400 °C cycling. All values are first-order estimates derived by simplified scaling of published empirical correlations [5–7, 9] from related coating systems; they should not be interpreted as exact predictions.

E. Planned AWJ Process Window Study

Prior to the full coating deposition and adhesion testing campaign, a preliminary AWJ process window study will be conducted to verify that the groove and taper geometries specified in the patent claims are achievable with available AWJ equipment. Trial runs will be performed on a CNC-controlled AWJ system at water pressures of 200–350 MPa with mesh 80 garnet abrasive. The objective will be to confirm that grooves within the specified depth range and width range can be reliably produced with acceptable dimensional repeatability. Based on published data [10, 11], the natural kerf fillet radius is expected to fall in the range of 60–90 μm. Tapered edge profiles with effective angles of 25–40° will be attempted by programming a linearly increasing traverse speed across the edge zone; some deviation from a perfectly linear profile is anticipated and will be quantified.

Surface profilometry of the inter-groove flat regions will be used to characterize the micro-roughness produced by AWJ processing, which is expected to yield Ra = 6–11 μm based on similar studies in the literature—comparable to conventional grit blasting [10]. The outcomes of this pre-study will serve to finalize the AWJ parameters for the subsequent full-scale specimen fabrication.

IV. PROPOSED NEXT STEPS

As the present work constitutes a conceptual and design framework rather than an experimental study, the forthcoming period will focus on experimental validation. The highest-priority task is the fabrication and tensile adhesion testing of all four specimen groups to obtain the first direct experimental evidence of the adhesion enhancement predicted by analysis. The measured adhesion values will be compared against the predictions in Table I to validate the theoretical framework.

The second task is thermal cycling durability testing, which is essential for demonstrating the edge taper benefit that is predicted to be most significant under cyclic loading. Specimens from all groups will be subjected to cyclic thermal loading (RT to 400°C, up to 500 cycles), with crack initiation and propagation monitored by periodic cross-sectional examination at intervals of 50, 100, 250, and 500 cycles. Based on the predictions (Table I), the integrated Group D specimens are expected to exhibit 2–3× longer crack initiation life compared to the baseline Group A.

The third task is groove geometry optimization through a focused factorial experiment varying groove depth (100, 150, 200 μm), width (250, 350, 500 μm), and pitch (500, 750, 1000 μm) to identify the combination yielding maximum adhesion for the specific APS alumina system under study. Additionally, a modified lap-shear test will be developed to measure interfacial shear strength, which is the loading mode where the groove mechanism provides the greatest theoretical benefit.

Several risks may reduce the predicted adhesion gains in practice. Incomplete splat filling of high-aspect-ratio grooves, AWJ-induced subsurface damage, and non-uniform coating thickness on steep groove walls are all potential concerns. Each could introduce defect populations that partially offset the mechanical interlocking benefit. As noted in Section III-D, discrepancies of ±30 % or more between predicted and measured values would not be unexpected. Additionally, the AWJ process window for simultaneously achieving the target groove geometry, taper angle, and micro-roughness may be narrow and require parameter compromises. These risks have been anticipated in the experimental plan. Mitigation measures include cross-sectional SEM inspection of groove filling and substrate integrity, systematic thickness mapping, and separate specimen groups that allow each modification to be evaluated independently.

REFERENCES

[1] L. Pawlowski, *The Science and Engineering of Thermal Spray Coatings*, 2nd ed. Chichester: John Wiley & Sons, 2008.

[2] P. Fauchais, J. O. Heberlein, and M. I. Boulos, *Thermal Spray Fundamentals: From Powder to Part*. New York: Springer, 2014.

[3] M. Amada, M. Tomitaka, and K. Hatanaka, "Effects of grit blasting conditions on the adhesion strength of plasma sprayed coatings," *Surf. Coat. Technol.*, vol. 102, no. 1–2, pp. 132–137, 1998.

[4] A. G. Evans and J. W. Hutchinson, "The mechanics of coating delamination in thermal gradients," *Surf. Coat. Technol.*, vol. 201, no. 18, pp. 7905–7916, 2007.

[5] S. H. Song, P. Xiao, and L. Q. Weng, "Numerical study of the effect of edge geometry on thermal fatigue of plasma sprayed coatings," *Surf. Coat. Technol.*, vol. 188–189, pp. 40–45, 2004.

[6] R. Kromer, S. Costil, J. Cormier, L. Berthe, P. Peyre, and D. Muot, "Laser patterning pretreatment before thermal spraying: A technique to adapt and control the surface topography," *J. Therm. Spray Technol.*, vol. 25, no. 3, pp. 401–410, 2016.

[7] K. Bobzin, T. Schläfer, M. Heeg, M. Öte, and R. Nicke, "Investigation of mechanical interlocking of thermal spray coatings on structured substrates," *J. Therm. Spray Technol.*, vol. 24, no. 7, pp. 1176–1184, 2015.

[8] D. B. Bogy, "Two edge-bonded elastic wedges of different materials and wedge angles under surface tractions," *J. Appl. Mech.*, vol. 38, no. 2, pp. 377–386, 1971.

[9] M. Arai, H. Nishida, and W. Lacarbonara, "Reduction of free-edge stress concentration in thermally sprayed coatings by edge shape optimization," *Eng. Fract. Mech.*, vol. 187, pp. 151–166, 2018.

[10] Y. Huang, H. Liang, and J. Wang, "Adhesion enhancement of thermal spray coatings with abrasive waterjet surface preparation," *J. Therm. Spray Technol.*, vol. 29, no. 5, pp. 1012–1025, 2020.

[11] N. Haghbin, J. K. Spelt, and M. Papini, "Abrasive waterjet milling of channels in metals: Model to predict high-aspect-ratio channel profiles," *J. Manuf. Process.*, vol. 17, pp. 11–2

Necessity of Sealing for Atmospheric Plasma Sprayed Alumina Insulation Coatings Against Humidity

[1]*Tomáš KMEC (5th year)*
Supervisor: [2]Juraj ĎURIŠIN

[1,2]Dept. of Technologies in Electronics, FEI, Technical University of Košice, Slovak Republic

[1]tomas.kmec@student.tuke.sk, [2]juraj.durisin@tuke.sk

***Abstract*— This paper demonstrates that atmospheric plasma sprayed (APS) Al_2O_3 coatings used for electrical insulation require post-spray sealing to maintain reliable dielectric performance under humid conditions. Alumina coatings (350 µm) were deposited on steel substrates using a GTV VIP 6000s plasma gun. Twelve specimens were divided into two groups of six: impregnated with a solvent-based sealant and non-impregnated. Insulation resistance was measured at 500 V DC using a KIKUSUI TOS9300. Results confirm that sealing is important for maintaining insulation integrity under humidity, although the significant degradation observed even in sealed coatings motivates future work on optimizing sealant type and impregnation method.**

***Keywords*— atmospheric plasma spraying; alumina coating; sealing; insulation resistance; climate testing; humidity; porosity; KIKUSUI TOS9300; dielectric degradation; impregnation.**

I. Introduction

Atmospheric plasma sprayed (APS) alumina (Al_2O_3) coatings are widely used as electrical insulation layers on metallic substrates in the electronics, energy, and aerospace industries [1]. However, a fundamental limitation of the APS process is that the resulting microstructure is inherently porous, with interconnected open porosity typically ranging from 3 % to 15 % [2]. When coated components are exposed to humid environments, water vapor penetrates the pore network through adsorption and capillary condensation, forming conductive pathways that severely degrade the insulation resistance [3]. The recognized solution is post-spray sealing—infiltrating a sealant into the porosity to block moisture ingress [4]. Despite broad acknowledgement that sealing is beneficial, there is a scarcity of published data directly comparing sealed and unsealed APS alumina coatings using standardized climate testing with a multi-stage measurement protocol. This study provides such a comparison, demonstrating that sealing important, and that further optimization of sealant selection and application methods is needed.

II. Preliminary investigation

The initial phase of this research established the measurement methodology, validated the instrumentation, and characterized the baseline behavior of the APS alumina coatings. All insulation resistance measurements were performed using a KIKUSUI TOS9300 Insulation Resistance Tester (Fig. 1), which provides a measurement range up to 50 GΩ, programmable test voltages, and compliance with IEC 61557-2 [5].

Fig. 1. KIKUSUI TOS9300

Preliminary measurements were performed on as-sprayed alumina-coated specimens at 250 V, 500 V, and 1000 V DC. The results confirmed voltage-independent behavior in this range (ohmic regime), and 500 V DC (current limitation 15 nA) was selected for all subsequent measurements as it provided the best signal-to-noise ratio and is consistent with industrial practice for coatings of this thickness class. The built-in charge/discharge circuitry ensures proper sample conditioning before measurement, and automated data logging facilitates systematic data collection. Repeatability was quantified by measuring a single specimen ten times: the coefficient of variation was 1.8% without connector repositioning, reflecting spatial variability in the coating microstructure.

Temporal stability was verified by measuring specimens daily over five consecutive days at room conditions (23 ± 1°C, 45 ± 3 % RH); the maximum deviation from Day 1 was 2.3 %. Direct terminal measurement configuration was adopted using soldered connectors on the specimens.. A 5-second stabilization period was used for all readings to allow decay of absorption and polarization currents [6]. Based on these preliminary results, a three-stage measurement protocol was defined: baseline measurement at ambient conditions, measurement immediately after climate exposure, and measurement after a five-day recovery period.

III. Experimental methodology

A. Sample preparation and Climate Test

Test specimens were prepared from a series-produced heating element (Fig. 2) and the directly soldered power connector was used for measurement.

Fig. 2. Heating element measurement

The NiCr bond coat thickness (~10 μm) was applied, corresponds to the specific production setup used in this study and is lower than conventional APS bond coat thicknesses (typically 50–150 μm). Followed by Al_2O_3 top coats deposited using the GTV VIP 6000s plasma gun with Al_2O_3 powder (D_{50} = 45 ± 6 μm, purity ≥ 99.7 wt. %). Target thickness was 350 ± 30 μm. A total of 12 specimens from the same spray session were divided into two groups: Group I (6 impregnated) and Group N (6 non-impregnated).

Sealing was performed by manual brush painting Group I specimens with a commercially available solvent-based impregnation sealant (WEICON S 500 HT) at room temperature under ambient pressure, followed by curing at 150 °C for 2 hours. The cured silicone fills interconnected pore channels with a hydrophobic, electrically insulating material stable to approximately 200 °C. This particular sealant was selected to demonstrate the principle of sealing effectiveness rather than to assert it as optimal for all applications.

The climate test was a steady-state damp heat exposure at 40°C/98 % RH for 16 hours per IEC 60068-2-78 [7]. Specimens were positioned vertically on ceramic standoffs in a Binder MKF 56 climate chamber.

B. Three-Stage Measurement Protocol

Stage 1 — Baseline (R_0): All specimens were conditioned at 23°C/45 % RH for 48 hours; insulation resistance was measured at 500 V DC with 5 s stabilization. **Stage 2 — Post-exposure (R_{exp}):** Measurements were completed within 5 minutes of removal from the climate chamber. **Stage 3 — Recovery (R_{rec}):** After five days at ambient conditions (23°C, ~45 % RH) with no active drying. Percentage degradation was calculated as:

$$\Delta Rexp = (Rexp - R0) / R0 \times 100\,\%$$

and recovery as:

$$Recovery = Rrec / R0 \times 100\%$$

C. Results

Table I presents the individual insulation resistance values for all twelve specimens at the three measurement stages. Table II provides a statistical summary.

The baseline resistance after production differed between groups: impregnated specimens averaged 11.58 GΩ while non-impregnated specimens averaged 9.83 GΩ. The initial difference in baseline insulation resistance between groups (~18%) may influence the comparative interpretation of the results and should be taken into account. However, no formal statistical hypothesis test (e.g., t-test or Mann–Whitney U test) was performed to determine whether this difference is statistically significant, and with only six specimens per group, the statistical power to detect moderate between-group differences is limited. Therefore, it cannot be conclusively determined whether the baseline difference reflects a genuine effect of the sealant on dry-condition resistance or simply specimen-to-specimen variability within the spray batch. For the same reason, the reported percentage differences between groups at subsequent measurement stages (post-exposure, recovery) should be interpreted as observed trends rather than statistically confirmed effects. This limitation applies to all between-group and within-group comparisons reported below. This suggests that the silicone sealant, by filling interconnected pores with a hydrophobic dielectric polymer, may improve the effective resistive path length even under dry conditions. However, this difference could also partly reflect specimen-to-specimen variability within the spray batch, and further investigation with larger sample sizes is warranted.

The contrast after climate exposure was dramatic. Non-impregnated specimens suffered a near-total loss of insulation resistance, dropping by 98.9 % from a mean of 9.83 GΩ to only 0.11 GΩ (110 MΩ). Every specimen in this group fell below 0.15 GΩ, with the worst case (sample N-4) reaching just 0.06 GΩ. For most practical insulation applications, such values would fall below minimum acceptable thresholds. In stark contrast, impregnated specimens decreased by 65.3 %, from 11.58 GΩ to 4.02 GΩ. While this is still a substantial degradation, the post-exposure insulation resistance of the impregnated group was 37 times higher than that of the non-impregnated group (4.02 vs. 0.11 GΩ). This order-of-magnitude difference underlines the critical protective role of the sealant.

After five days of recovery at ambient conditions, non-impregnated specimens recovered to a mean of 7.30 GΩ, corresponding to 74.3 % of baseline. This dramatic rebound from near-zero values demonstrates that the moisture-induced degradation mechanism is largely reversible: as water desorbs from the pore network, conductive pathways are progressively eliminated. Nevertheless, 25.7 % of the original resistance remained unrecovered, indicating that moisture trapped in deep, fine porosity requires substantially longer to desorb. Impregnated specimens recovered to 8.97 GΩ (77.5 % of baseline)—a slightly higher recovery ratio, consistent with the expectation that less moisture entered the sealed microstructure and therefore less remains to desorb.

TABLE I
Individual insulation resistance measurements (GΩ) at 500 V DC, 5 s stabilization

Stage - Impregnated	**I-1**	**I-2**	**I-3**	**I-4**	**I-5**	**I-6**
After production (R_0)	12.4	11.8	11.9	11.1	11.0	11.3
After climate (R_{exp})	4.6	4.0	4.2	3.8	3.7	3.8
After climate +5 d (R_{rec})	9.5	8.3	8.4	8.8	8.9	9.9
Stage - Non-Impregnated	**N-1**	**N-2**	**N-3**	**N-4**	**N-5**	**N-6**
After production (R_0)	10.2	9.8	9.4	10.8	9.5	9.3
After climate (R_{exp})	0.12	0.15	0.08	0.06	0.14	0.11
After climate +5 d (R_{rec})	7.1	6.9	7.4	7.9	7.3	7.2

IV. DISCUSSION

The 98.9 % degradation of non-impregnated coatings confirms the extreme vulnerability of as-sprayed APS alumina to humidity. At 98 % RH and 40°C, capillary condensation fills pores with effective radii below the Kelvin critical radius (~5–10 nm at these conditions), converting gas-filled insulating voids to liquid-water-filled conductive pathways throughout the coating thickness [3]. The interconnected porosity characteristic of the APS lamellar microstructure allows moisture to bridge the full coating cross-section, creating a near-short-circuit condition. The silicone sealant partially disrupts this mechanism by occupying pore channels with a hydrophobic polymer that resists capillary condensation and blocks water transport.

However, the 65.3 % degradation of impregnated specimens reveals that ambient-pressure silicone impregnation does not achieve complete pore sealing. The finest inter-splat boundaries and the deepest pore networks likely remain unfilled, allowing partial moisture penetration. This is a significant finding: while sealing transforms the outcome from electrical short-circuit failure (0.11 GΩ) to a degraded but potentially serviceable state (4.02 GΩ), there remains substantial room for improvement through better sealants, vacuum-assisted impregnation, or multi-step sealing processes.

The recovery behavior provides additional insight. The fact that non-impregnated specimens recovered from 0.11 to 7.30 GΩ (a 66-fold increase) within five days indicates that the dominant degradation mechanism is physical moisture presence rather than permanent chemical or microstructural damage. If electrochemical corrosion or irreversible phase transformation had occurred, recovery would be far less complete. The incomplete recovery of both groups (74–78 % of baseline) after five days suggests that full desorption of moisture from the APS microstructure requires either extended time or active drying (e.g., elevated temperature bake-out).

From a practical standpoint, the results lead to an unambiguous conclusion: APS alumina coatings in the as-sprayed, non-impregnated condition must not be specified for electrical insulation in any application where humidity exposure is anticipated, unless the design explicitly accounts for a potential two-order-of-magnitude reduction in insulation resistance. Silicone impregnation provides important protection, maintaining resistance at gigaohm levels, but the 65 % degradation indicates that current sealing practice is not yet optimized and that application-specific sealant selection is essential.

Several limitations of this study should be noted. Only one sealant type and one impregnation method (ambient-pressure immersion) were evaluated. Only a single climate profile was tested; more severe profiles (e.g., 85°C/85 % RH) may challenge the sealant further. Long-term durability under repeated humidity cycling was not assessed. No microstructural characterization (SEM, mercury intrusion porosimetry) was performed to quantify the degree of pore filling. These limitations define the directions for future work.

V. PROPOSED NEXT STEPS

Having established that sealing is necessary—and that the current silicone impregnation, while highly beneficial, still permits significant degradation—the principal direction of future work is the systematic identification of the optimal sealant type and sealing method. The following tasks are planned for the forthcoming period.

First, a comparative evaluation of multiple sealant categories will be conducted using the same three-stage protocol established here. Candidate sealants include solvent-based impregnation (the current baseline), epoxy resin, sol–gel derived alumina and silica infiltration, phosphate-based ceramic sealers (aluminum phosphate, chromium phosphate), and polyurethane impregnation. Each will be applied to specimens from the same spray batch for direct comparison.

Second, the influence of application method will be investigated. Vacuum-assisted impregnation, in which the specimen is first evacuated and then immersed while vacuum is released, is expected to achieve deeper and more complete pore filling than the ambient-pressure immersion used in this study, and quantifying this improvement is critical for reducing the 65 % degradation observed here [4].

Third, thermal stability will be assessed by subjecting sealed specimens to thermal aging at 200–500°C prior to climate testing. This is neccesary for high-temperature applications (motor windings, heating elements) where organic sealants may decompose and inorganic alternatives (sol–gel, phosphate) are required [8]. Fourth, long-term durability under repeated humidity cycling (100+ cycles) will be evaluated. Fifth, microstructural characterization by SEM/EDS and mercury intrusion porosimetry will correlate the degree of pore filling with insulation performance. Finally, a guarded three-terminal electrode configuration per IEC 62631-3 will be implemented to separate surface leakage from volume conduction [6]. The practical outcome of this program will be a decision framework mapping application requirements to the recommended sealant and impregnation method.

TABLE II
SUMMARY STATISTICS FOR IMPREGNATED VS. NON-IMPREGNATED GROUPS

Group Impregnated	R_0 (GΩ)	R_{exp} (GΩ)	ΔR_{exp} (%)	R_{rec} GΩ)	REC (%)
Impregnated	11.58	4.02	-65.3	8.97	77.5
Non-Impregnated	9.83	0.11	-98.9	7.3	74.3

REFERENCES

[1] L. Pawlowski, *The Science and Engineering of Thermal Spray Coatings*, 2nd ed. Chichester: John Wiley & Sons, 2008.

[2] P. Fauchais, G. Montavon, and G. Bertrand, "From powders to thermally sprayed coatings," *J. Therm. Spray Technol.*, vol. 19, no. 1–2, pp. 56–80, 2010.

[3] F.-L. Toma, L. M. Berger, D. Jacquet, D. Pecheur, T. Mäntylä, and T. Luhman, "Comparative study on the electrical insulation properties of thermally sprayed Al_2O_3 coatings," *Surf. Coat. Technol.*, vol. 185, no. 2–3, pp. 232–240, 2004.

[4] J. Knuuttila, P. Sorsa, T. Mäntylä, J. Kiilakoski, and P. Vuoristo, "Sealing of thermal spray coatings by impregnation," *J. Therm. Spray Technol.*, vol. 8, no. 4, pp. 599–604, 1999.

[5] KIKUSUI Electronics Corp., "TOS9300/TOS9301 Insulation Resistance Tester — Instruction Manual," Yokohama, Japan, 2018.

[6] IEC 62631-3-1:2016, *Dielectric and resistive properties of solid insulating materials — Part 3-1: Determination of resistive properties (DC methods)*, International Electrotechnical Commission, Geneva, 2016.

[7] IEC 60068-2-78:2012, *Environmental testing — Part 2-78: Tests — Test Cab: Damp heat, steady state*, International Electrotechnical Commission, Geneva, 2012.

[8] D. Leivo, T. Varis, E. Turunen, M. Vippola, T. Mäntylä, and P. Vuoristo, "Sealed alumina coatings for high-temperature applications," *J. Therm. Spray Technol.*, vol. 18, no. 3, pp. 445–451, 2009.

Author's index

B
Badár Jozef 27
Balara Viliam 65
Basarik Tomáš 151
Bobček Marek 244
Bubeňková Lenka 120
Buček Tomáš 148

Č
Čatloch Dušan 33

D
Danko Tomáš 17
Demeter Peter 81
Dopiriak Matúš 114

E
Eliáš Kristián 50
Eliáš Martin 68

F
Fodor Samuel 144

G
Glajc Kristián 220
Gordan Daniel 154

H
Herasymchuk Nazarii 137
Horváth Marek 165
Hresko Sebastian 200
Hric Ladislav 21
Hric Matej 211
Hyseni Ardian 183

Ch
Chovancová Simona 93

I
Imrich Miroslav 194

J
Jusko Miroslav 134

K
Katonová Erika Abigail 226
Kmec Tomáš 253, 256, 259
Kmecik Tadeáš 217
Kohútová Monika 127
Komorník Miroslav 234
Kormaník Tomáš 238
Kováč Juraj 77
Kováčová Antónia 205
Král Richard 208
Krempaský Alex 179
Kupcova Eva 36
Kurimský František 176

L
Leľová Michaela 157
Liščinská Zuzana 74
Líšková Dominika 214

M
Mahdal Filip 123
Margita František 24
Marko Jaroslav 104
Maťašová Sylvia 97
Molnár Roland 161
Murin Miroslav 141

N
Nemergut Peter 117
Nguyen Martin 131
Nováková Ivana 111
Novotný Samuel 241

O
Onufer Samuel 172

P
Pallaj Ľubomír 186
Podracký Erik 12
Poprik Peter 197

R
Rešovský Norbert 248
Rusnáková Renáta 30

S
Setlák Branislav 190
Sobota Timotej 107
Sopkovic Kristian 223
Stromko Michal 54
Syzranova Svitlana 85
Szomosi Vladimir 58

Š
Ševc Kamil 231
Ševec Jerguš 39
Šimčák Július 62
Šofranko Peter 168
Šolc Daniel 43

T
Tomaščík Lukáš 71
Tomaščíková Nikola 47

U
Uličný Chris 89

V
Vaľko Matej 100

SCYR 2026: 26th Scientific Conference of Young Researchers

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 264 pages
Available Online

Editors:
Assoc. Prof. Ing. Karol Kyslan, PhD.
Assoc. Prof. Ing. Emília Pietriková, PhD.
Ing. Lukáš Pancurák, PhD.

ISBN 978-80-553-5051-6