Scientific Conference of Young Researchers 2021

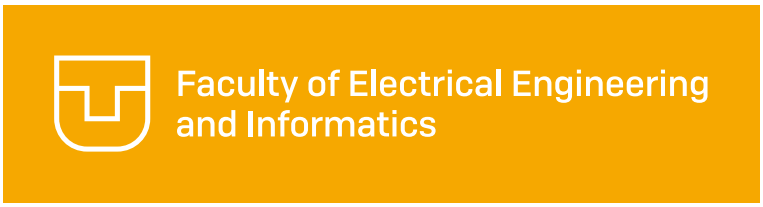
21st Scientific Conference of Young Researchers

April 9, 2021
Košice, Slovakia

Proceedings from Conference

Faculty of Electrical Engineering and Informatics
Technical University of Košice

Sponsors


Faculty of Electrical Engineering
and Informatics


SES
člen
ZSVTS


elfa


SIEMENS
Healthineers

SCYR 2021: 21st Scientific Conference of Young Researchers
Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 202 pages, number of CD Proceedings: 50 pieces

Editors: prof. Ing. Alena Pietriková, CSc.
Ing. Emília Pietriková, PhD.

ISBN 978-80-553-3904-7

Scientific Committee of SCYR 2021

General chair:	Prof. Ing. Liberios Vokorokos, PhD.
Editorial board chairman:	Prof. Ing. Alena Pietriková, CSc.
Proceeding reviewers:	Prof. Ing. Roman Cimbala, PhD. Prof. Ing. Iraida Kolcunová, PhD. Prof. Ing. Ján Paralič, PhD. Prof. Ing. Daniela Perduková, PhD. Prof. Ing. Alena Pietriková, CSc. Prof. Ing. Jaroslav Porubän, PhD. assoc. Prof. Ing. František Babič, PhD. assoc. Prof. Ing. Ján Gamec, CSc. assoc. Prof. Ing. Ján Genči, PhD. assoc. Prof. Ing. Kristína Machová, PhD. assoc. Prof. Ing. Ján Papaj, PhD. assoc. Prof. RNDr. Jana Tóthová, PhD.

Organizing Committee of SCYR 2021

Members:	Prof. Ing. Alena Pietriková, CSc. Ing. Ivana Olšiaková Ing. Emília Pietriková, PhD.
Contact address:	Faculty of Electrical Engineering and Informatics Technical University of Košice Letná 9 042 00 Košice Slovak Republic

Foreword

Dear Colleagues,

SCYR (Scientific Conference of Young Researchers) is a Scientific Event focused on exchange of information among young scientists from Faculty of Electrical Engineering and Informatics at the Technical University of Košice – series of annual events that was founded in 2000. Since 2000, the conference has been hosted by FEEI TUKE with rising technical level and unique multicultural atmosphere. The 21^{st} Scientific Conference of Young Researchers (SCYR 2021) was held on April 9, 2021.Due to COVID-19 pandemics, the conference was held online.The primary aims of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities at the Faculty of Electrical Engineering and Informatics, has been achieved. Approx. 70 participants, mostly by doctoral categories, were active in the conference.

Faculty of Electrical Engineering and Informatics has a long tradition of students participating in skilled labor where they have to apply their theoretical knowledge. SCYR is an opportunity for doctoral and graduating students to train their scientific knowledge exchange. Nevertheless, the original goal is still to represent a forum for the exchange of information between young scientists from academic communities on topics related to their experimental and theoretical works in the very wide spread field of a wide spectrum of scientific disciplines like informatics sciences and computer networks, cybernetics and intelligent systems, electrical and electric power engineering and electronics.

Traditionally, contributions can be divided in 2 categories:

- Electrical & Electronics Engineering
- Computer Science

with approx. 70 technical papers dealing with research results obtained mainly in university environment. This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the Faculty of Electrical Engineering and Informatics. This Scientific Network included various research problems and education, communication between young scientists and students, between students and professors. Conference was also a platform for student exchange and a potential starting point for scientific cooperation. The results presented in papers demonstrated that the investigations being conducted by young scientists are making a valuable contribution to the fulfillment of the tasks set for science and technology at the Faculty of Electrical Engineering and Informatics at the Technical University of Košice.

We want to thank all participants for contributing to these proceedings with their high quality manuscripts. We hope that conference constitutes a platform for a continual dialogue among young scientists.

It is our pleasure and honor to express our gratitude to our sponsors and to all friends, colleagues and committee members who contributed with their ideas, discussions, and sedulous hard work to the success of this event. We also want to thank our session chairs for their cooperation and dedication throughout the entire conference.

inally, we want to thank all the attendees of the conference for fruitful discussions and a pleasant stay in our event.

Liberios VOKOROKOS
Dean of FEI TUKE

April 9, 2021, Košice

Contents

Matej Madeja
Discovering Testing Practices in GitHub Open-Source Projects 10

Ján Magyar
Supporting Personalization with Reinforcement Learning 12

Jakub Palša
The Procedure Steps for Malware Analysis using Machine Learning 14

Peter Havran
Investigation of the condition of insulating oil through mathematical transformations in impedance spectroscopy 16

Rastislav Petija
Anomaly detection in the Internet of Things environment based on data flows 18

Daniel Pál
The role of electric vehicles in smart grids 21

Róbert Štefko
Protection system for microgrid and smart grid systems 23

Anastázia Margitová
Transient-State Temperature Calculation of Overhead Power Line Conductors 27

Vladimír Kohan
Monitoring of power system using WAMS technology – during disturbances 30

Marek Ružička
Synthetic End-user Trajectory Generation 32

Marek Fedor
Dynamic systems control of asynchronous motor drive 35

Marek Kuzmiak
Superconducting properties of strongly disordered molybdenum nitride ultrathin films in the vicinity of superconductor-insulator transition 39

Fouzia Adjailia
Robust 3D Facial Expression Recognition 42

Maroš Lapčák
Environmental impacts on hybrid FSO/RF systems 44

Miloš Šárpataky
Dielectric nanofluids facing sustainable development 48

Ľuboš Šárpataky
Diagnostics methods of insulators in wet and polluted conditions 54

Máté Hireš
A survey on deep learning based pathological speech detection 59

Peter Bugata
Feature selection based on neural networks 61

Branislav Fecko
System for data collection from a CNC milling machine for the purpose of making a thermal model 63

Dávid Martinko
Impact of Photovoltaic systems and electric vehicles on a residential low voltage power grid 65

Marián Hudák
Multi-user adaptive interface for web-based collaborative virtual environments 67

Jakub Urbanský
Utilization of battery energy storage systems in frequency regulation 69

Norbert Zdravecký
Available knowledge of 64x100 Gb/s DWDM system with different modulations 71

Stanislav Alexovič
An overview of Autonomous Mobile Robot research and multi cooperation 75

Simeon Samuhel
Magnetic properties of amorphous ferromagnetic glass coated microwires 79

Juraj Biľanský
Design of Battery Tester 83

Jozef Ivan
The Test Bench for Testing High Precision Actuators 85

Július Bačkai
Geometrically Frustrated Magnetic Systems under High Pressure 87

Daniel Gecášek
A general framework for scientific computing 91

Matej Gazda
Convolutional neural networks for chest x-ray medical imaging 93

Jozef Humeník
Digitization and transformation of the electricity market 95

Michal Kolárik
Explainability of artificial intelligence in medical domain 97

Zuzana Pella
Analyzing and identifying potential factors most significantly affecting CVD 102

Viera Pavlišková
Intelligent Methods Using Data to Support Decision Making 105

Ľudmila Pusztová
Rules Adaptation Mechanism in the Medical Diagnostic Process 109

Martin Havrilla
Characterization of Computer Flow Data Based on Computer Learning to Streamline Network Traffic 111

Martin Hasin
Analysis and Evaluation of Data Using Artificial Intelligence for Cybersecurity 115

Slavomír Gereg
Deep neural networks for speech-to-text systems 119

Daniel Dzivý
Influence of Multiple Reflows on Electrical Resistivity of Solder Joints 121

Maksym Oliinyk
Research on the impact of electric vehicles and renewable energy sources on the electrical grid . 123

Maksym Karpets
Characterization and SANS study of transformer oil-based magnetic fluids 125

Michal Solanik
Methods and algorithms for acceleration / parallelization of physical model calculations 127

Martin Sivý
Controlling SMART devices via unified user interface in Virtual Reality environment 131

Maroš Baumgartner
Robust data transmission in 5G networks without infrastructure 133

Viera Maslej Krešňáková
Overview: Text Classification using Deep Learning Models 137

Filip Gurbáľ
Improving Program Comprehension Using Testing Methods and Tools 139

Anton Buday
Application of Transformer neural architecture in speech recognition of Slovak language 143

Michal Hulič
Cryptography accelerator using symmetric and asymmetric algorithms 145

Viktor Petro
Position-Sensorless Control of Permanent Magnet Synchronous Motor 147

Marianna Koctúrová
Speech activity detection with visual stimuli 151

Lukáš Hruška
Enhancing reinforcement learning agents with a priori knowledge using deictic representation .. 153

Ladislav Pomšár
Acceleration of AI on the edge 155

Ivan Čík
Study of visualization methods of explainable artificial intelligence 157

Adrián Marcinek
Review of Multiport Converters Topologies 160

Michal Varga
Contribution to Generative Enhancement of 3D Image Classifiers 165

Jana Horniaková
Domain wall dynamics in rapid changing magnetic field 167

Tomáš Lenger
Thermal Properties of Substrates for Face-down Technology 169

Peter Hrabovský
The GUI design in MATLAB environment for optimization parameters of spiral planar coil ... 171

Andrinandrasana David Rasamoelina
Importance of Normalization in Meta-Learning ... 173

Milan Tkáčik
Modeling and Implementation of Distributed Control Systems for Large Physical Experiments .. 176

Miroslav Sokol
Application specific integrated circuits for UWB sensors systems ... 178

Lukáš Koska
Final Proposal on Modeling and Control for Walking Robots ... 181

Richard Olexa
Simulation Model of a Small Hydropower Plant ... 183

Martin Durkáč
Overview of Modeling and Simulation Methods for Mechanical Systems ... 186

Peter Šuľaj
Image stitching based on drone technologies ... 190

Renát Haluška
Hybrid FSO/RF Systems and Cloud ... 192

Rikin Tailor
Smart Grid and Microgrid in an Industrial Zone ... 194

Miroslav Jaščur
Lung and Pleura Segmentation In Lung Ultrasound with Deep Learning ... 199

Author's Index ... 201

Discovering Testing Practices in GitHub Open-Source Projects

[1]*Matej MADEJA (4th year),*
Supervisor: [2]*Jaroslav PORUBÄN*

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

[1]matej.madeja@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—Software testing is one of the very important Quality Assurance (QA) components. There are many instructions on how to write tests correctly, but it is unknown how tests are written in practice. To analyze such tests it is necessary to effectively find them in a project. In this paper it was investigated whether the "test" word correlates with the presence of test cases. The analysis was performed on 38 GitHub open source repositories thoroughly selected from the set of 4.3M GitHub projects with the Java as majority language. We analyzed 20,340 test cases in 803 classes manually and 170k classes using an automated approach. The results show that there exists weak correlation ($r = 0.655$) between the word *test* and test cases presence in a class. Simultaneously, an automated script was proposed using static file analysis with success of 95% correctly detected test cases.**

***Keywords*—Testing practices, program comprehension, test case identification, searching test by name, GitHub.**

I. Introduction

Many authors [1], [2], [3] define best practices for tests to facilitate comprehension of the testing code and the identification of test failure. Some guidelines lead to avoiding test smells [4] because as reported by recent studies [5], [6], their presence might not only negatively affect the comprehension of test suites but can also lead to test cases being less effective in finding bugs in the production code. All mentioned approaches are only recommendations, but do not really express how the tests are written in real projects. That means we know how tests should be written, but we do not know how they are written in practice. Many researchers have tried to clarify the motivation of writing tests [7], [8], [9], the impact of test-driven development (TDD) on code quality [10], [11] or the popularity of testing frameworks [12]. According to our knowledge, there is no research that deals with the analysis of tests in independent projects.

To reveal testing practices in real and independent projects it is necessary to find a way to identify test cases in a project, without the time-consuming code analysis. Much more important than the number of test cases is the information where they are located. When a testing framework is used, the test identification is mostly straightforward, e.g. by the presence of the framework imports. On the other hand, it is advisable to consider tests that do not use any third-party framework and can be regarded as customized testing solutions. Considering the ways in which test cases were identified in related work and based on the authors' experience of Java test cases development, it can be assumed that there is a relation between the word *test* and the number of test cases in a file. That means searching for the *test* string could be beneficial for faster test case identification. Based on the previous reasoning, we focused on finding the correlation between the number of occurrences of the word test in the project and real numbers of test cases.

II. Method

GitHub was used as data source of analysed projects. The set of projects was obtained via GHTorrent [13] due to its high availability. The experiment was limited to projects whose the most used language is Java. In existing research is common technique searching for testing frameworks' imports [14] or to search for files with the world *test* in the filename [15].

Instead of searching "test" string we could run test via automated build tool (e.g. ant, maven or gradle) and collect the number of tests. In that case, the issue is that building such open source projects often fail [16] and we need to build every single project and run tests what is a time consuming task. Therefore, we tried to count and especially find the location of such test cases without compilation.

Based on the meaning of the word *test* we assumed that there will be a correlation between the occurrence of the word *test* (in file content or filename) and the number of test cases. Therefore, three datasets were created using the searching GitHub API[1] for:

1) the word *test* in filename,
2) the word *test* in file content,
3) frameworks' imports in file content (38 collected frameworks).

The word "test" was searched for every single project in 3 mentioned places, 4.3 million projects in total. It is possible to expect that the more occurrences of the word *test* in the project, the more test cases will be present in it. Therefore, 38 projects[2] with the highest occurrence of the word test (in file content or filename) or with the highest occurrence of a specific framework's import were selected for manual analysis, during which an automated script[3] was developed. A script was created to automate the identification of test cases in the future. Searching for *test* regardless of the framework, we were also able to analyze testing practices without using

[1]https://docs.github.com/en/rest
[2]See more at https://github.com/madeja/unit-testing-practices-in-java.
[3]https://github.com/madeja/unit-testing-practices-in-java/blob/master/AnalyzeProjectCommand.php

any third-party framework, i.e. customized testing solutions. Because GitHub contains many projects that are not relevant, e.g. testing, homework or cloned projects, rules for searching relevant projects have been defined, resulting in set of projects used for the analysis.

III. Results

Using the automated script all 38 repositories' files were processed, 38 repositories and 170,076 classes altogether, from which 803 classes and 20,340 test methods were manually investigated. To evaluate the precision of the proposed automated script results were compared to manual test identification of 20,340 test cases across all three datasets. Accuracy of 95.72% for test cases detection was achieved by automated identification. Most of false positives and false negatives occurrences were caused by customized testing solutions, e.g. when tests were performed directly from the `main()` function by calling methods of the class. If the naming conventions of the called (testing) methods were not governed by the principles of frameworks (e.g. prepending method name with *"test"* or using public methods), not all test cases were detected in an automated way.

Figure 1 shows the correlation with the linear regression line of the word "test" and the number of test cases in particular class. A standard (pearson) correlation coefficient of $r = 0.655$ was reached, that is not strongly significant when considering significance level $\alpha = 0.05$. However, from the perspective of finding projects containing tests, this technique is beneficial and can help future experimenters to filter projects containing tests much more faster, without need of build. Because projects had different numbers of test classes and used different frameworks, the detailed ratio of the word "test" occurence and test case presence per project can be found at GitHub[4].

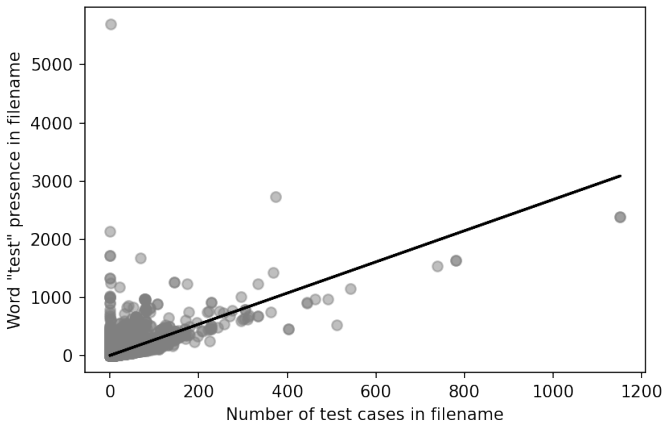


Fig. 1. Correlation of the word "test" presence and number of test cases for analyzed classes by automated script.

IV. Future Directions

As future work, we plan to find a solution for an accurate identification of test cases in customized solutions. We believe that by studying testing practices, it will be possible to train artificial intelligence to automatically recognize tests by the structure and nature of the code. At the same time, we would like to focus on mining tests for information that could support the production source code comprehension and streamline the development process.

[4]https://github.com/madeja/unit-testing-practices-in-java/blob/master/correlation-boxplot.png

Acknowledgment

This work was supported by project VEGA No. 1/0762/19: Interactive pattern-driven language development.

References

[1] A. Nayyar, *Instant Approach to Software Testing: Principles, Applications, Techniques, and Practices*. BPB Publications, 2019.

[2] W. E. Lewis, *Software testing and continuous quality improvement*. CRC press, 2017.

[3] B. Garcia, *Mastering Software Testing with JUnit 5: Comprehensive guide to develop high quality Java applications*. Packt Publishing Ltd, 2017.

[4] A. Van Deursen, L. Moonen, A. Van Den Bergh, and G. Kok, "Refactoring test code," in *Proceedings of the 2nd international conference on extreme programming and flexible processes in software engineering (XP)*, 2001, pp. 92–95.

[5] A. Peruma, K. Almalki, C. D. Newman, M. W. Mkaouer, A. Ouni, and F. Palomba, "On the distribution of test smells in open source android applications: an exploratory study." in *CASCON*, 2019, pp. 193–202.

[6] D. Spadini, F. Palomba, A. Zaidman, M. Bruntink, and A. Bacchelli, "On the relation of test smells to software code quality," in *2018 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 2018, pp. 1–12.

[7] M. Linares-Vásquez, C. Bernal-Cardenas, K. Moran, and D. Poshyvanyk, "How do developers test android applications?" in *2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 2017, pp. 613–622.

[8] M. Beller, G. Gousios, A. Panichella, and A. Zaidman, "When, how, and why developers (do not) test in their ides," in *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, ser. ESEC/FSE 2015. New York, NY, USA: Association for Computing Machinery, 2015, p. 179–190. [Online]. Available: https://doi.org/10.1145/2786805.2786843

[9] P. S. Kochhar, F. Thung, N. Nagappan, T. Zimmermann, and D. Lo, "Understanding the test automation culture of app developers," in *2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST)*, 2015, pp. 1–10.

[10] D. Fucci, H. Erdogmus, B. Turhan, M. Oivo, and N. Juristo, "A dissection of the test-driven development process: Does it really matter to test-first or to test-last?" *IEEE Transactions on Software Engineering*, vol. 43, no. 7, pp. 597–614, 2017.

[11] W. Bissi, A. G. Serra Seca Neto, and M. C. F. P. Emer, "The effects of test driven development on internal quality, external quality and productivity: A systematic review," *Information and Software Technology*, vol. 74, pp. 45 – 54, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950584916300222

[12] A. Zerouali and T. Mens, "Analyzing the evolution of testing library usage in open source java projects," in *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2017, pp. 417–421.

[13] G. Gousios, "The ghtorrent dataset and tool suite," in *Proceedings of the 10th Working Conference on Mining Software Repositories*, ser. MSR '13. Piscataway, NJ, USA: IEEE Press, 2013, pp. 233–236. [Online]. Available: http://dl.acm.org/citation.cfm?id=2487085.2487132

[14] P. Stefan, V. Horky, L. Bulej, and P. Tuma, "Unit testing performance in java projects: Are we there yet?" in *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering*, ser. ICPE '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 401–412. [Online]. Available: https://doi.org/10.1145/3030207.3030226

[15] P. Leitner and C.-P. Bezemer, "An exploratory study of the state of practice of performance testing in java-based open source projects," in *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering*, ser. ICPE '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 373–384. [Online]. Available: https://doi.org/10.1145/3030207.3030213

[16] M. Sulír, M. Bačíková, M. Madeja, S. Chodarev, and J. Juhár, "Large-scale dataset of local java software build results," *Data*, vol. 5, no. 3, p. 86, 2020.

Supporting Personalization with Reinforcement Learning

[1]*Ján MAGYAR (4th year),*
Supervisor: [2]*Peter SINČÁK*

[1,2]Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]jan.magyar@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—Intelligent systems often adjust their behavior to the needs of their users through the use of reinforcement learning, but successful application is hindered by a combination of poor sample efficiency and a lack of training data. We show that neural network architectures tailored to the personalization problem can mitigate some negative effects on a test scenario having the characteristics of personalization problems while abstracting away from their complexity. Our solution shows comparable or better sample efficiency than existing methods, while leading to a higher level of personalization and better overall policies.**

***Keywords*—personalization, Q-networks, reinforcement learning**

I. Introduction

As digital systems become an ever more important part of everyday human life, it is desirable that they show behavior tailored to the needs and preferences of individual users. This process is called *personalization*, whose goal is to change "the functionality, interface, information access and content, or distinctiveness of a system to increase its personal relevance to an individual or a category of individuals" [1].

One of the drawbacks of current personalization approaches is that they only adjust an already defined functionality of a system, which still requires that this basic functionality be hard-coded. In some cases, the correct system behavior is hard to define, or we would prefer if the system were to find it autonomously using methods of artificial intelligence.

In this paper we present some results of our research into using a neural network architecture specifically designed to enable a system to learn a primary functionality and personalize it at the same time. Section II provides a quick overview of the state of the art in personalization with reinforcement learning, Section III describes our architecture, Section IV presents the methodology of our evaluation, the results of which are included in Section V with Section VI concluding the paper.

II. Personalization and Reinforcement Learning

Reinforcement learning (RL) describes the learning problem in the context of an interaction between a learning agent and an environment, which makes it highly adaptable to the problem of user interaction personalization. In this context, the digital system, or a component of it, constitutes the learning agent, while the interaction itself with the user's preferences represents the environment. The goal of the system is then to get the environment to a desired state through careful and deliberate action selection.

In [2], the authors provide a systematic overview of examples of using RL for personalization, and define three main types of approaches. In the first one, each user is represented by a separate environment and the system is equipped with separate learning agents for each user. This approach does not create an overhead for personalization, but experience transfer is impossible. In the second approach, description of the user becomes a part of the environment's state, which enables experience transfer, but it is still hard to find the user attributes most relevant to the personalization. The third approach tries to balance the advantages and disadvantages of the previous two approaches in that it personalizes system behavior to a group of users. Experience transfer is possible within a group, but individual agents must be trained for each group. Additionally, defining a function mapping users to their respective groups is another challenge.

As shown in [3], the application of RL for personalization might be hindered by poor sample efficiency. As most methods of machine learning, RL often requires a large amount of training data to learn the desired behavior, which is hard to amass in the context of human–computer interaction: we would like the system to personalize its behavior to the user's preferences in as few interaction steps as possible.

III. Q-Network with Dynamically Loaded Biases

While reinforcement learning is often used to adjust existing system behavior, in many use cases it is not straightforward what the system functionality should be. It is possible to define the desired goal of the interaction, but the steps that need to be taken to achieve this goal are unclear. Defining a single system behavior might lead to monotonous interactions and would reflect the bias and presumptions of the system designer.

We therefore propose to use Q-networks to learn the primary system behavior as well as personalize it to preferences of individual users. To this end (see Fig. 1), we define two types of input: S is a vector representing the state of the environment, while V is a vector identifying the user or group of users through one-hot encoding. This means that in vector V, apart from a single 1 value, all input values are 0. Hidden and output units are then connected to all units in the previous layer, and all input units of input V. Hidden and output units have no biases of their own. Since we know that V corresponds to a one-hot encoded identificator of an environment, we can see that weights connecting input V to later layers serve as biases that are loaded dynamically based on the environment

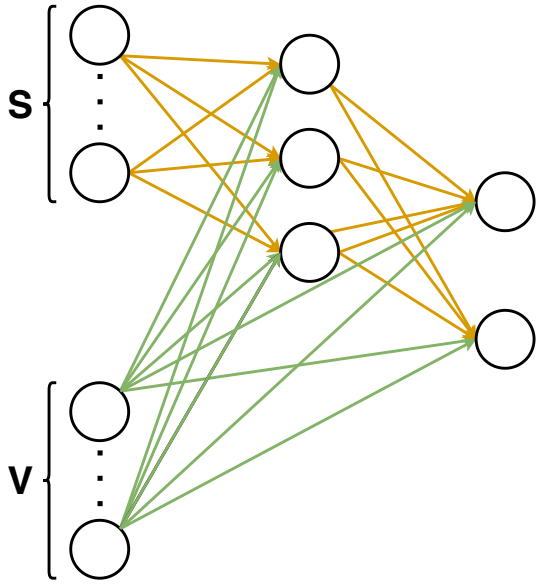


Fig. 1: Dynamically Loaded Biases Q-Network architecture

the agent is interacting with. This lends the architecture its name – Dynamically Loaded Biases Q-Network (DLBQN).

Unlike table-based Q-learning, Q-networks can diverge, an issue that can be addressed through stabilization methods. Our solution uses two main stabilization methods, *experience replay* [4] and *separate target network* [5]. To better support experience transfer and sample efficiency, we introduce *federated learning* into the learning algorithm [6]. We tested two variants of the algorithm: in which federated learning was used to train the primary network – `FedDLBQN`, and in which federated learning was used to train the target network – `TargetFedDLBQN`.

IV. Methodology

We tested our network in a scenario that has the characteristics of personalization problems while it abstracts away from their complexity, the Push The Box environment. In this task, the agent must learn to cooperate with a simulated human to push a box to a certain goal position of a 2D environment along the shortest way. The box's position is updated only if the human and the agent push it in the same direction. We defined three types of humans: 1) pushing the box first to the same column as the goal position, then to the same row; 2) pushing the box vertically first, then horizontally; 3) pushing the box along the dimension where the distance is higher (or horizontally). A reward of 1 was observed upon reaching the goal position, and a punishment of -1 was given for all additional steps.

We trained Q-networks with three hidden layers with 24, 48, and 24 nodes respectively. As a benchmark, we considered a simple Q-network corresponding to the second approach described in Section II (`SwitchQN`): the network had the same inputs, but V was only connected to the first hidden layer, and nodes had biases. We ran 500 test runs on eight 3×3 worlds with randomly generated goal positions and a maximum of 100 training iterations per world. Training was stopped once the agent found the optimal policy – learned to replicate the humans' actions for all positions in all worlds. We used two types of state representation: box position only (2 parameters) and goal position with box position (4 parameters). We considered the average number of iterations needed to find the optimal policy to evaluate sample efficiency, and root-mean square distance of predicted Q-values to calculate the level of personalization.

state	method	mean iterations	mean precision	mean RMS
2	SwitchQN	**153.55**	99.98	0.48
	FedDLBQN	163.22	**100**	**0.57**
	TargetFedDLBQN	159.83	**100**	0.56
4	SwitchQN	148.43	99.96	0.31
	FedDLBQN	152.61	**100**	**0.37**
	TargetFedDLBQN	**146.21**	**100**	0.36

TABLE I: Experiments results

V. Results

The results of our experiments are shown in Table I. As can be seen, only the proposed `FedDLBQN` and `TargetFedDLBQN` methods were able to find the optimal policy for all test runs. With two-parameter state representation, `SwitchQN` had a slightly better sample efficiency, but the difference is smaller than 1 iteration/world. `FedDLBQN` showed slightly more personalized behavior when compared to `TargetFedDLBQN`. Both approaches resulted in higher level of personalization than `SwitchQN`.

VI. Conclusion

In this paper we described our method, Dynamically Loaded Biases Q-Network, which addresses some issues characteristic to personalization problems. Compared to existing methods, our approach shows a comparable sample efficiency, but a higher level of personalization. Additionally, while existing methods only adjust a system's existing behavior, our method can be used to find that primary behavior as well. Extended results of our research have been published at the 19th World Symposium on Applied Machine Intelligence and Informatics and are in the review process to a special issue of IEEE Transactions on Human-Machine Systems. In future research, we will look for ways to ensure better sample efficiency and test our approach in a real-life personalization setting.

Acknowledgment

This research is supported by AI4EU project from the European Union's Horizon 2020 research and innovation programme under grant agreement 825619 2019-2021, and the Maria Curie RISE LIFEBOTS Exchange Grant Agreement ID : 824047 2019-2021.

References

[1] J. Blom, "Personalization: a taxonomy," in *CHI'00 extended abstracts on Human factors in computing systems*, 2000, pp. 313–314.

[2] F. den Hengst, E. M. Grua, A. el Hassouni, and M. Hoogendoorn, "Reinforcement learning for personalization: A systematic literature review," *Data Science*, no. Preprint, pp. 1–41, 2020.

[3] J. Magyar, M. Kobayashi, S. Nishio, P. Sinčák, and H. Ishiguro, "Autonomous robotic dialogue system with reinforcement learning for elderlies with dementia," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 3416–3421.

[4] L.-J. Lin, "Reinforcement learning for robots using neural networks," Carnegie-Mellon Univ Pittsburgh PA School of Computer Science, Tech. Rep., 1993.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[6] C. Nadiger, A. Kumar, and S. Abdelhak, "Federated reinforcement learning for fast personalization," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*. IEEE, 2019, pp. 123–127.

The Procedure Steps for Malware Analysis using Machine Learning

[1]*Jakub PALŠA (2nd year),*
Supervisor: [2]*Liberios VOKOROKOS*

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

[1]jakub.palsa@tuke.sk, [2]liberios.vokorokos@tuke.sk

***Abstract*—The theme of this dissertation is malware detection based on static and dynamic analysis of machine learning. The work details the topics of malware, static analysis, dynamic analysis as well as the technology of machine learning, which has enriched us with knowledge necessary for research in this area. The goal of previous work has been to process available harmful samples into a form which allows analysis and the preparation of safe environment for realizing predictions of static and dynamic analysis of malware samples. The next step is the realization of the prediction of static and dynamic analysis itself using machine learning algorithms (Decision Tree, Random Forest, Naive Bayes, and Support Vector Machine). After the prediction, the static and dynamic analysis algorithms will be compared, and the success of the results evaluated for research purposes.**

***Keywords*—Machine learning, Static analysis, Dynamic analysis, Malware.**

I. Introduction

Today, we already live in the age of predictive marketing, autonomous driving, facial recognition, and personalized digital assistants using machine learning – artificial intelligence. The extent to which these inventions make our lives easier, increase our security or threaten our privacy is up to our judgement, but we can say for sure that the rate of cybercrime will not decrease even due to these technologies, rather the opposite [1]. This invites the question, why not use these technologies to fight against cybercrime?

The goal of this work is to suggest a model of classification of machine learning predictions, which will be able to sort samples into safe and dangerous categories. The emphasis will be placed on minimizing mistakes where the system labels a dangerous sample as a safe one, because we consider minimizing real vectors of attack a priority and consider the side effect of false identification of safe samples as dangerous to be the lesser security risk.

This work could potentially bring new light on the subject of malware identification, new services of companies which are oriented on fighting cyber-attacks or at least expand the set of data available for dynamic analysis which are available to be used again to come up with a new solution, which can help combat these cybernetic threats.

II. Solution Proposal

This chapter describes the steps made to obtain the successful results necessary for the research of this dissertation. It describes the environment for testing, used test samples, data processing and prediction step by step. The specification of the order of the problem-solving process is displayed in picture *1*.

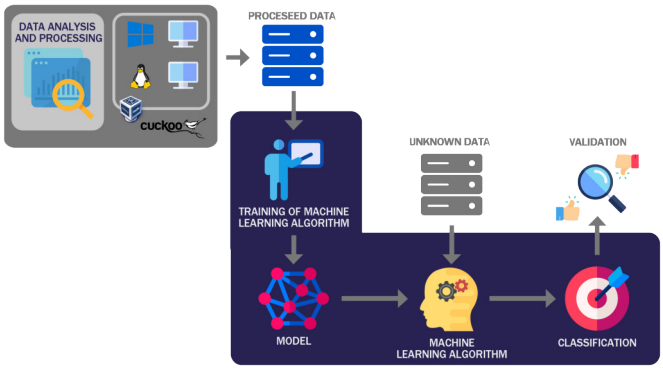


Fig. 1. The sequence of steps of the solution process

A. Environment Preparation

We chose Microsoft Windows 7 as a suitable environment for static and dynamic analysis for two reasons [2]:

1) The greatest sample of malware available to us is up to date for this version of the OS.
2) Newer versions of Windows OS include better security systems and the activation of the malware would not be guaranteed.

After installing and activating the OS, we installed applications on it which have nothing to do with the solution itself and whose only purpose is to create an environment similar to a computer of a common user. For the same reason we did not install VirtualBox Guest Additions, which will lower the performance and possibilities of the virtualization, but also lower the risk of detection by malware.

B. Data Collection

The basic input for the prediction model is data. Considering that the task was classifying an unknown sample into two categories (clean sample, harmful sample), it was necessary to acquire a sufficiently large amount of clean and harmful samples. We decided to study the executable files on Windows because these files are attacked most often. We succeeded in creating a set of approximately 2000 clean samples and about 5000 harmful samples. The set of harmful samples was acquired from an online repository *virusshare.com* [3], which

provides a collection of many harmful samples for malware research and analysis. Because not all these samples were executable files, we had to write a script in Python which sorted these samples and collected only executables.

After collecting the necessary samples, we needed to collect information about every single sample. For this task, we chose the freely available tools Dependency Walker and Robotask for static analysis and Cuckoo Sandbox for dynamic analysis [4]. After analyzing the samples using these tools, we saved the acquired data in a text file. It was unthinkable to do so manually, considering that 7000 samples had to be analyzed, which is why we utilized above mentioned tools to do so.

C. Data Processing

The acquired text files contained a huge amount of data on the samples. We only wanted to focus on the important details, based on which we would be able to differentiate a clean sample from a harmful one. Publicly available Windows API gives malware creators huge possibilities [5], which is exactly what we wanted to focus on. The book [6] describes which of the functions are most often used by malware.

For outcome prediction we have used Anaconda and Python in version 3.6. The Anaconda package presents a toolbox for Data Science, which is the use of scientific procedures for processing data in order to make conclusions. From a scientific perspective, this package provides tools for quick work with different kinds of data, the possibility of quick environment switching and the use of prediction tools without previous experience.

Scikit-learn

Scikit-learn provides a library for Python, which aims to facilitate the tasks of mining and analyzing data. Not only does this package provide tools for classification, regression, or clustering, but also the tools for preprocessing and dimensionality reduction of data.

III. Following Steps

For predicting the results it necessary to find suitable parameters, using which the algorithm returns the best results overall. The procedure for static analysis will be differentiated from dynamic analysis.

A. Static Analysis

Static analysis [7] will be done using DependencyWalker. The generation of files will be realized using Robotask. Which function was used is an interesting part of the generated data, which is why a final *.csv* file will be created, containing:

- a column for each function used,
- a row for each malware analyzed.

Subsequently, this *.csv* file will be used as the data set for chosen algorithms.

B. Dynamic Analysis

The dynamic analysis [8] of malware samples consists of activating them and observing their behavior and interaction with the environment. To lower the risk of infecting real systems, this will be done in a secure virtual environment. Executing the samples is the responsibility of Cuckoo Sandbox. The sample can be uploaded through a web interface or using the command line. After uploading, the sample is activated inside a virtual environment prepared in advance. The execution of malicious code and its effects on the system are observed for a certain predetermined period.

Cuckoo Sandbox can analyze malware and record information such as:

- the system calls of all malware created processes,
- created, modified, and deleted files,
- network activity, and
- memory registry.

All this information is saved into an analysis message in .json format. Information about system calls will be used for the purposes of this work. The records of all malware samples will be processed into a table to create a dataset for machine learning. The table will contain:

- columns representing the names of system functions,
- rows representing single malwares analyzed and
- cells at the column and row intersection representing the number of calls for a given function for a given sample.

The purpose of the data set is to train and test the qualifier.

C. Results Comparison

After a successful prediction of all chosen qualifiers, it is possible to continue to analysis and interpretation of results. The easiest way of rating qualifier success is to determine its classification accuracy. Accuracy is given as the ratio of correctly classified samples to the ratio of incorrectly classified samples.

IV. Conclusion

In this paper we describe the steps already determined and the steps which we will undertake to successfully realize our research. Based on our research a security tool able to detect unknown samples of harmful software might be developed in the future. Using different models or creating a new one might be an interesting expansion of the solution which has not been included in this paper.

Acknowledgment

This work was supported by the Faculty of Electrical Engineering and Informatics at the Technical University of Košice under contract No. FEI-2021-76: A modern interpreter of predicate linear logic formulas.

References

[1] J. Kolouch, *CyberCrime*, 2016.
[2] K. Monnappa, "Learning malware analysis: Explore the concepts, tools, and techniques to analyze and investigate windows malware," *Packt Publishing Ltd*, 2018.
[3] VirusShare.com, "Virusshare." [Online]. Available: https://virusshare.com
[4] M. e. a. Vasilescu, "Practical malware analysis based on sandboxing," *RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference*, pp. 1–6, 2014.
[5] H. e. a. Tamada, "Dynamic software birthmarks based on api calls," *IEICE Transactions on Information and Systems*, 2016.
[6] M. e. a. Sikorski, *Practical malware analysis: the hands-on guide to dissecting malicious software*, 2012.
[7] A. e. a. Moser, "Limits of static analysis for malware detection," *Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007)*, pp. 421–430, 2007.
[8] U. e. a. Bayer, "Improving the efficiency of dynamic malware analysis," *In Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 1871–1878, 2010.

Investigation of the condition of insulating oil through mathematical transformations in impedance spectroscopy

[1]*Peter HAVRAN (2nd year)*
Supervisor: [2]Roman CIMBALA

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

[1]peter.havran@tuke.sk, [2]roman.cimbala@tuke.sk

***Abstract*—This article deals with the investigation of Shell DIALA S4 ZX-1 insulating oil subjected to accelerated thermal aging. The first part contains the measured data in the time domain of impedance spectroscopy, from which in the second part the data were subsequently converted to the frequency domain, using two mathematical transformations with reference to dielectric losses of insulating oil in the low-frequency band.**

***Keywords*—complex susceptibility, frequency domain, impedance spectroscopy, time domain.**

I. Introduction

Electrical insulating oils are one of the main components of the insulation system of power equipment, including transformers, as the most important equipment in the power system. It is necessary to use reliable methods of examination and diagnostics to eliminate problems caused by equipment failure during operation, which could cause technical and economic damage [1].

This experiment is aimed at measuring and investigating the world-unique inhibited insulating oil Shell DIALA S4 ZX-1 (SD), produced in Qatar based on GTL (gas to liquid) technology. This technology involves the process of converting natural gas into liquid hydrocarbons, which results in pure insulating oil, which is characterized by high oxidative stability, zero sulfur content, and excellent thermal properties. SD insulating oil was diagnosed by impedance spectroscopy (IS) [2].

The essence of the non-destructive IS method lies in the analysis of dielectric relaxation to reduce dielectric losses in materials that reflect the condition of insulation to prevent damage related to the failure of electrical equipment [1].

II. Investigation of the Condition of Insulating Oil

The continuity of the transformer power depends to a large extent on the quality of the insulating oil. Thermal degradation and chemical reactions such as oxidation reduce the performance of transformer oil, deteriorating its insulating and cooling properties. The degradation of transformer oil is affected by the high temperature at which the oil is loaded by accelerated thermal aging [5].

A. Experimental measurements

Experimental measurements were performed in the time domain IS on a sample of insulating oil SD, which was stressed by accelerated thermal aging at a temperature of 90 °C for 1830 hours. Measurements at 0 h were not performed due to the high moisture content in the oil, as it is not known how the oil was stored before. Several measurements have been made, but in this publication, only the first measurement (after 250 h) and the last measurement (after 1830 h) of charging currents at 1000 s at an applied voltage of 40 V will be pointed out. The experiment contained professional measuring components in the form of Keysight 16452A electrode system, KEITHLEY 6517B meter, laptop, and UF55^{plus} universal oven, which ensured accelerated thermal aging.

Fig. 1 shows a graphical dependence of the charging current on time at different times of thermal aging of oil SD, where it is possible to see an increased current in the form of capacitive, polarization, and conductivity current at accelerated thermal aging of 1830 h.

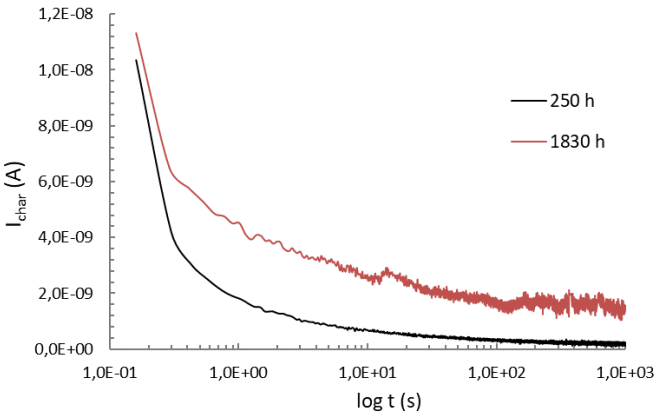


Fig. 1. Measured data in the time domain of impedance spectroscopy

B. Mathematical transformations

The frequency domain of impedance spectroscopy (IS) was performed through mathematical transformations from the time domain to the frequency domain. These mathematical conversions were performed using the Fourier transform and Hamon approximation, with their subsequent comparison. Fig. 2 shows the graphical dependence of the imaginary part of the complex susceptibility on the frequency at different times of thermal aging through both mathematical transformations.

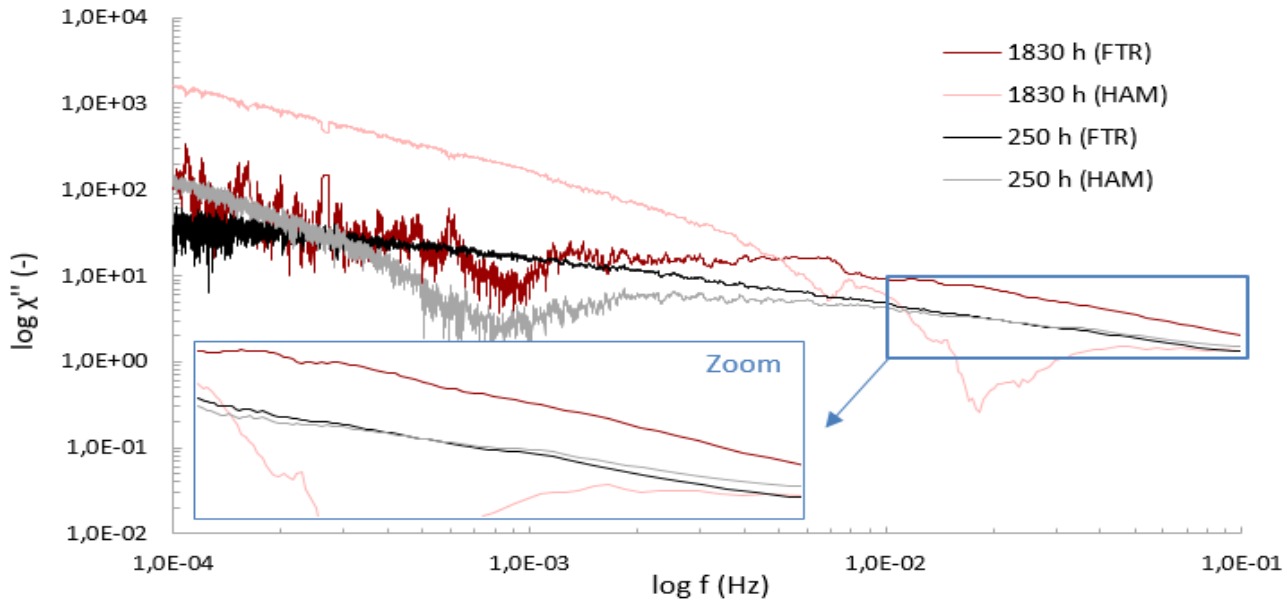


Fig. 2. Calculated data in the frequency domain of impedance spectroscopy

The imaginary part of the complex susceptibility χ'' was obtained using the Fourier transform from the dielectric response function φ(t), which is closely related to the charging current [3]:

$$\chi^*(\omega) = \chi'(\omega) - j\chi''(\omega) = \int_0^\infty \phi(t)\, e^{-j\omega t} dt \qquad (1)$$

where ω is the angular frequency and t is time. The second transformation method was the Hamon approximation, by which the imaginary part of the complex susceptibility was obtained from the dielectric response function via the formula:

$$\chi''(\omega) = \frac{|-\phi(t) C_0 U|}{\omega C_0 U} - \frac{\sigma_0}{\varepsilon_0 \omega} \qquad (2)$$

where C_0 is the no-load capacitance, U is the voltage, σ_0 is the unidirectional conductivity and ε_0 is the permittivity of the vacuum [4].

From the graphical dependencies in Fig. 2, it is possible to see increased values of the imaginary part of the complex susceptibility in the low-frequency band after thermal aging of 1830 h. It means, there was an increase in the dielectric losses of the SD oil similar to that shown in Fig. 1 via charging currents. From the comparison of transformation methods, it can be stated that the Fourier transform achieves more accurate results than the Hamon approximation, whose behavior is similar to a certain oscillation around the values of the Fourier transform. Therefore, the Fourier transform, whether its direct or inverse function, will be used in preference to the Hamon approximation in further measurements and investigations of insulating materials, or possible conversions between time and frequency domain in impedance spectroscopy.

III. Conclusion and Further Research

This publication provides a graphical interpretation of dielectric losses in SD insulating oil using two transformation methods in impedance spectroscopy. Further research will focus on the temperature characteristics of SD oil and magnetic nanofluid from SD insulating oil, respectively.

Acknowledgment

The work was supported by the Ministry of Education, Science, Research and Sport within the project VEGA 1/0154/21 and 2/0011/20 and the Slovak Agency for Research and Development based on contracts no. APVV-18-0160 and APVV-15-0438.

References

[1] P. Havran, R. Cimbala, P. Bartko, "Accelerated Thermal Aging of the Liquid Dielectrics," 2019 International IEEE Conference and Workshop in Óduba on Electrical and Power Engineering (CANDO-EPE), pp. 207-212, ISBN: 978-1-7281-4358-3.

[2] R. Cimbala, P. Havran, P. Bartko, "Comparison of behavior of electroinsulating oils using thermo-hysteresis dependencies," 2020 21st International Scientific Conference on Electric Power Engineering (EPE), pp. 1-5, ISBN: 978-1-7281-9479-0.

[3] D.K. Das-Gupta, P.C.N. Scarpa, "Polarization and dielectric behavior of ac-aged polyethylene," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 3, no. 3, 1996, ISSN: 1070-9878.

[4] M. Farahani, H. Borsi, E. Gockenbach, "Dielectric spectroscopy in time and frequency domain on insulation system of high voltage rotating machines," Proceedings of the 2004 IEEE International Conference on Solid Dielectrics, 2004. ICSD 2004, pp. 1-4, ISBN: 0-7803-8348-6.

[5] I.Y. Negara et al., "Analysis of Physical and Electrical Characteristics of Transformer Oil Insulation During Accelerated Thermal Aging Experiment and Its Lifetime Estimation Using Arrhenius Law and Breakdown Voltage Test," 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), pp. 139-143, ISBN: 978-1-5386-2708-2.

Anomaly detection in the Internet of Things environment based on data flows

[1]*Rastislav PETIJA (3rd year),*
Supervisor: [2]*František JAKAB*

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

[1]rastislav.petija@tuke.sk, [2]frantisek.jakab@tuke.sk

***Abstract*—In this work we will focus on the possibility of classifying data flows, designing an architecture for monitoring the quality of the IoT environment and the detection of anomalies, which most often include the detection of attacks. To solve the mentioned problems, the TinyIPFIX / IPFIX protocol will be used to transfer information about data flows from the Internet of Things environment to the controller. Subsequently, the analysis of data flows using machine learning will be performed on the controllers.**

***Keywords*—Internet of Things, Anomaly detection, TinyIPFIX, Network monitoring**

I. Introduction

Initial research in the Internet of Things was related to exploring the possibilities of data transmission, the origin of which was formed by intelligent devices such as various sensor systems that measure environmental characteristics such as temperature, pressure, humidity, etc. Due to the increasing number of IoT devices, the network infrastructures themselves and the quality requirements in such infrastructures have also started to increase. Current researches are beginning to deal with the classification of transmitted data as stated in [1] and also the need to create systems that can assess the quality of the IoT transmission network.

Before it is possible to start detecting anomalies, it is necessary to classify the observed data. Several procedures have been developed for classifying data, including Port-based that was described in [2], [3] and [4], Payload-based [5] and Behavioural-based approaches. For our purposes, where we want to work efficiently using the smallest possible resources, the most appropriate approach to classification is Statistical-based, that according to the authors in [6], use only statistical data of data transmissions, without the need for their in-depth control, which reduces the requirements for computing resources and speeds up the classification. Examples of statistical data are data such as duration of communication, size of transmitted packets, frequency of sending messages, etc.

II. Proposal of monitoring architecture

Part of this work was to design a comprehensive solution that would allow monitoring the environment of the Internet of Things. At the same time, it should be able to implement simple integration and cooperation with existing monitoring systems. The key components of the proposed system are mainly the IoT gateway and the collector. The monitoring of the IoT environment will be based mainly on the monitoring

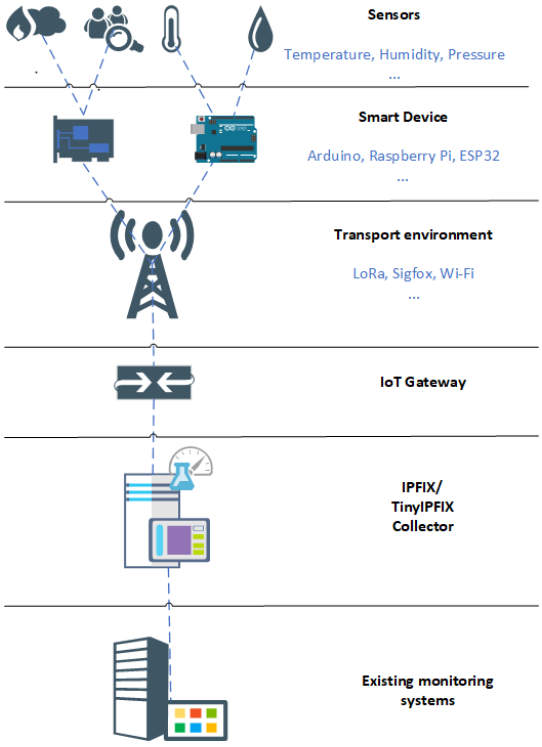


Fig. 1. Architecture of monitoring system

of the qualitative characteristics of data streams, on the transmission of which the TinyIPFIX protocol (data transmission in a limited IoT environment) and the IPFIX protocol (data transmission over a regular network) will be used. The architecture of the proposed system is shown in Fig. 1. The design also included a proposal for the functionalities of two main elements, namely the IoT gateway and the IPFIX/TinyIPFIX collector.

The IoT gateway has the following functionalities:

1) Receive data from IoT environment via technologies such as LoRa, WiFi
2) Selection of observed transmission parameters - depending on the transmission technology
3) Creation of a data model - selection of data intended for export
4) Data Export Process in IPFIX / TinyIPFIX format

The IPFIX/TinyIPFIX collector has the following functionalities:

1) Ability to collect data in IPFIX / TinyIPFIX format

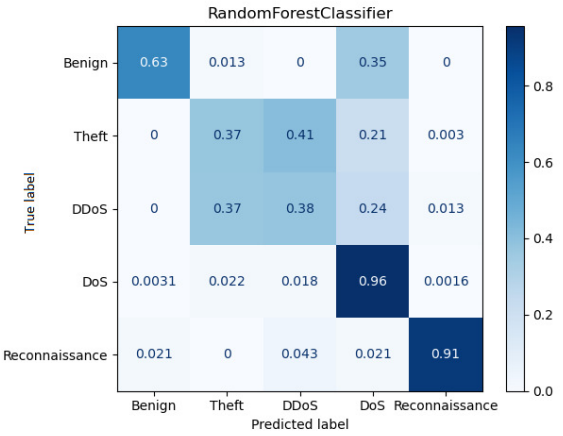


Fig. 2. Attack detection using Random Forest model - 1

2) Data analysis by statistical models of machine learning
3) Detection of infrastructure anomalies
4) Mediation process enabling conversion between TinyIPFIX and IPFIX protocol
5) The process of exporting data in IPFIX format, which will allow connect the monitoring of IoT environment with existing monitoring tools in common network infrastructures

The proposed architecture was also implemented and made it possible to perform the following experiment, which made it possible to verify the possibility of the proposed monitoring system to perform attack detection.

III. Anomaly detection based on IPFIX data

Due to the demanding process of creating a data set describing the real environment, publicly available data sets were used in our experiment. Anomaly detection was performed based on dataset, which was published in [7]. In this case, an anomaly is considered to be a situation where an attack is transmitted over the network. The data contained in the dataset were transferred by the TinyIPFIX protocol from the IoT device to the collector, where machine learning models were subsequently analyzed. The use of the KNN and Random Forest algorithms proved to be the most suitable for this purpose.

Datasets were created from a pcap file from which information about data streams was identified. We worked with two datasets. The first dataset, called NF-ToN-IoT, consisted of 1,379,274 data streams, of which 80.4% of data traffic was attacks and 19.6% of traffic was normal traffic. The dataset contained the following attacks: Backdoor, DoS, DDoS, Injection, MITM, Password, Ransomware, Scanning and XSS. The following confusion matrix shows the success of the detection of the used machine learning model, while in our case the best results were shown by the Random forest algorithm, which can be seen on Fig. 2. The second dataset, called NF-BoT-IoT, consisted of 600,100 data streams, of which 97.69% of data traffic was attacks and 2.31% of traffic was normal traffic. The mentioned dataset contained the following attacks: Reconnaissance, Theft, DoS and DDoS. The following confusion matrix shows the success of the detection of the used machine learning model, while in our case the best results were shown by the Random forest algorithm, which can be seen on Fig. 3.

This experiment was performed to confirm the possibility

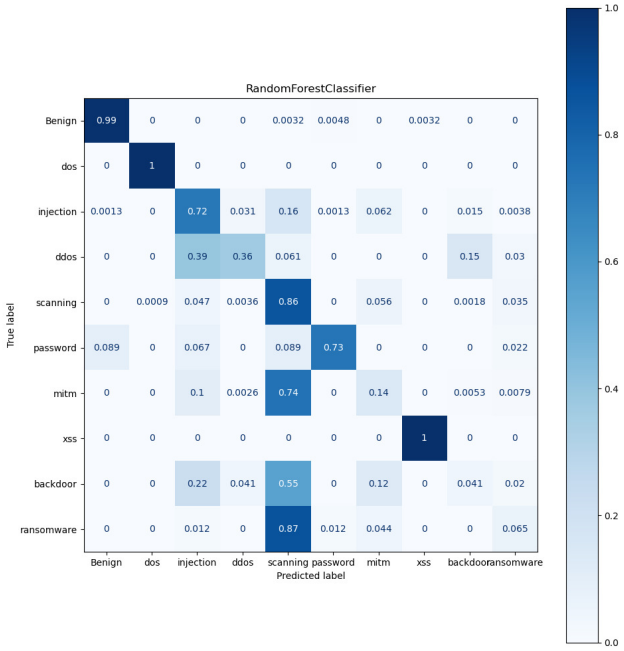


Fig. 3. Attack detection using Random Forest model - 2

of using the IPFIX and TinyIPFIX protocols to transmit all important information about IoT data streams through the environment thanks to exporter and collector that was implemented within this work. Another goal was to point out that based on the information about data flows obtained by the IPFIX protocol, it is possible to detect different types of attacks and at the same time to show which types of attacks are reliably detectable by the model created by us. In our case, based on the used datasets, it turned out that the proposed model is able to reliably detect the Reconnaissance, DoS and XSS attacks.

IV. Conclusion

It has been shown that the TinyIPFIX and IPFIX protocols can be used as protocols for the transmission of information about data streams also in the IoT environment. The created collector was able to capture and store this data in a form that was then easily used to detect attacks. Based on the positive results from the experiment, it is possible to continue to expand the functionality of the created system.

Further work will focus on capturing, transmitting and storing information about other anomalies, which may include detection of changes in data transmission frequency, loss of transmitted signal strength, monitoring of interference intensity and detection of changes in transmitted data content in IoT environment.

ACKNOWLEDGMENT

This paper has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF.

References

[1] H. Tahaei, F. Afifi, A. Asemi, F. Zaki, and N. B. Anuar, "The rise of traffic classification in IoT networks: A survey," *Journal of Network and Computer Applications*, vol. 154, p. 102538, mar 2020.

[2] G. D. L. T. Parra, P. Rad, and K.-K. R. Choo, "Implementation of deep packet inspection in smart grids and industrial internet of things: Challenges and opportunities," *Journal of Network and Computer Applications*, vol. 135, pp. 32–46, jun 2019.

[3] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos, "Is p2p dying or just hiding?" in *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04.* IEEE.

[4] A. W. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 41–54.

[5] N. A. Khater and R. E. Overill, "Network traffic classification techniques and challenges," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*. IEEE, oct 2015.

[6] J. Zhang, X. Chen, Y. Xiang, W. Zhou, and J. Wu, "Robust network traffic classification," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1257–1270, aug 2015.

[7] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "Netflow datasets for machine learning-based network intrusion detection systems." [Online]. Available: https://arxiv.org/abs/2011.09144

The role of electric vehicles in smart grids

[1]*Daniel PÁL (3rd year)*
Supervisor: [2]Ľubomír BEŇA

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

[1]daniel.pal@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract* — This publication examines the role of electric vehicles in smart grids. Electric vehicles, along with renewable energy sources, are becoming increasingly popular in the world. The capacity of renewables is constantly increasing, and the demand for electric cars is also showing an increasing trend. Electric vehicles, along with renewable energy sources play a key role in transforming today's grid into a smart grid. The first part of the publication describes smart grids. The second part deals with electric cars and shows how the demand for them has increased and is expected to increase in the coming years. The last part also deals with the connection of electric cars to the grid.**

***Keywords* — electric vehicles, Smart Grid, Vehicle-to-Grid (V2G).**

I. INTRODUCTION

Electric cars are becoming more and more popular not only in the European Union, but also in the world. The number of electric cars sold is increasing year by year. When from the traditional grid becomes an intelligent grid, electric cars and renewable energy sources will play a key role in grid security. Electric cars can also be able to solve the biggest disadvantage of renewables (especially solar panels) that they do not produce at night. If properly implemented in the grid, it is possible that electric cars will be charged by people when they are at work as the peak of solar panel production is during the day If they go home, these cars could serve the network. If they go home, these cars could serve the network. Their batteries could be used for network security.

II. SMART GRID

Defining a smart grid (shown in Fig. 1) in general is quite problematic because each state, each continent has a different approach to this issue. According to the European Commission, a Smart Grid can be described using the following considerations:

- flexibility - responds to consumer requirements,
- availability - all new sources, including renewable energy sources, can be connected to the network,
- reliability - always ensures the safety and quality of electricity supply,
- economy - efficient energy management [1].

According to the US Office of Electricity, an intelligent network is possibly characterized by the following aspects:

- higher efficiency in electricity transmission, lower losses,
- reduction of operating and management costs - reduction of energy prices for consumers,
- security - network resilience to physical or cyber interference,
- integration of renewable energy sources [2].

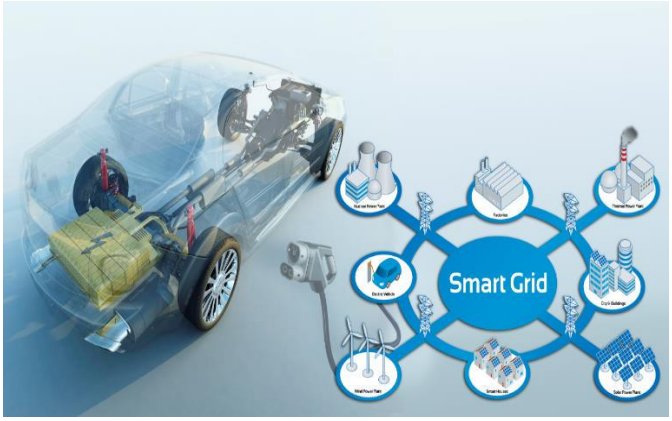


Fig. 1 Smart Grid [3]

III. ELECTRIC VEHICLES

As was shown in the Fig. 1, one of the key elements of smart grids will be electric cars. Electric cars are becoming more and more popular today than in the European Union, and the demand for them is growing all over the world. The number of electric cars is increasing year by year, which can also be seen in the following figure. In 5 years, their number has increased from 1.40 million to almost 8 million, which corresponds to a sixfold increase. Of this growth, China has the largest role, as China's electric car stock was nearly 4 million by the end of 2019.

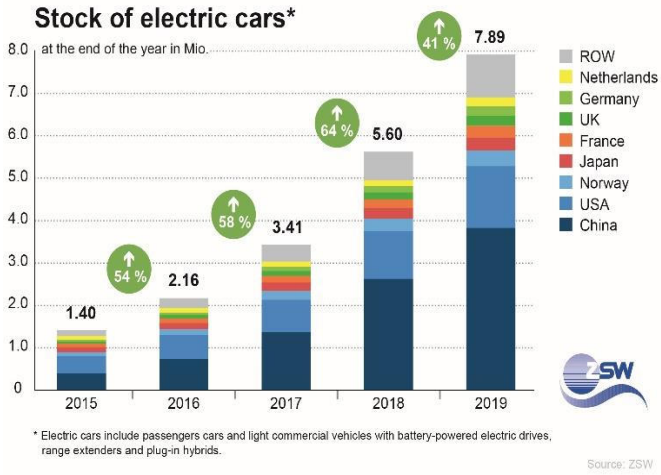


Fig. 2 Global electric car stock [4]

It is likely that the demand for electric cars will not decrease but will show an increasing trend. This can also be seen in the following picture that the demand for electric cars will continue to grow year by year. The interesting thing about the graph is that the world's car stock will be owned by 6 countries and the rest will be shared by the rest of the world. These larger countries are USA, China, Japan, Germany, United Kingdom and France. Of these, China has a bigger role based on the graph, as it is expected that more electric cars will be used there.

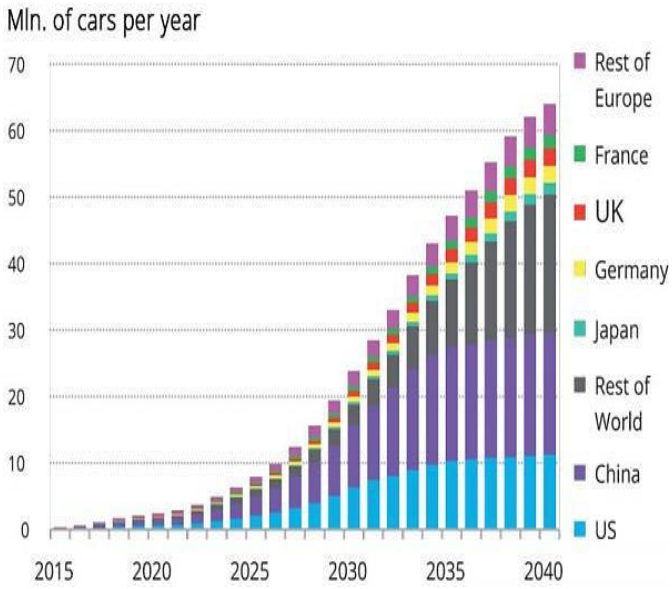


Fig. 3 Expected growth of electric cars [5]

IV. The Role of Electric Cars in Smart Networks

Electric cars can affect the smart grid in different ways depending on how we look at them. They can be:

- Consumers - when electric cars are connected to the grid to charge the battery.
- Producers - when the stored energy in the battery is returned to the grid.
- Batteries - when the excess energy that can be found in the network is stored for later use or the stored energy is returned to the grid. It is a combination of the previous two.

In smart grids, communication is two-way. This allows supply and demand to be matched at a local level, allowing for a flexible network. The vehicle-to-grid (V2G) system shown in the following figure can help with this. The conclusion is that a given electric car that is connected to the grid not only consumes power but can also send power back to the grid.

This allows in smart grids, for example, to store the surplus energy that is in the grid in car batteries and then use it when it is needed. This would solve the biggest disadvantage of solar panels. Photovoltaic panels use sunlight as a source of energy and generate direct current electricity, but these panels do not produce electricity at night. However, with the help of electric cars, the energy produced during the day would also be stored and then used at night if necessary.

In smart grids, the storage capacity of electric cars can solve another network problem. V2G-capable vehicle offers reactive power support, active power control, variable renewable energy monitoring, load balancing because the cars are capable of locally produced or consumed electricity. With the help of these can improve grid efficiency, stability, and reliability.

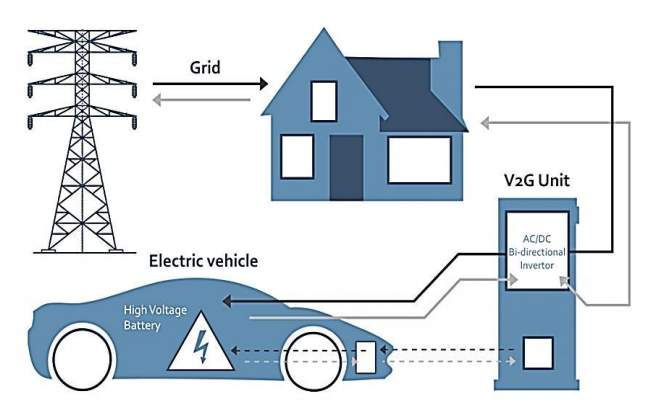


Fig. 4 Vehicle-to-Grid (V2G) system [6]

V. Conclusion

The aim of the publication was to draw attention to the possible impact of electric vehicles to the smart grid. It was showed that more and more electric cars are being sold, which is not expected to decrease, but even increase in the coming years. Therefore, it is very important to properly integrate into the network due to network security. Smart grids can help with this.

VI. Future Research

Further research focuses on setting up a computer program designed to calculate network losses based on input parameters. The loss-based program should determine the location of the distributed generations, renewable energy sources and electric vehicles to reduce the losses and increase network stability.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576 "Electricity self-sufficiency in conditions of liberalized electricity market".

References

[1] European SmartGrids Technology Platform – Vision and Strategy for Europe's Electricity Networks of the Future [Online]. Available at: <https://ec.europa.eu/research/energy/pdf/smartgrids_en.pdf> Accessed on February 1, 2021.

[2] The Smart Grid [Online]. Available at: <https://www.smartgrid.gov/the_smart_grid/smart_grid.html> Accessed on February 1, 2021.

[3] Inauguration of the European Interoperability Centre for Electric Vehicles and Smart Grids [Online]. Available at: <https://ec.europa.eu/jrc/en/event/other-event/inauguration-interoperability-centre> Accessed on February 1, 2021.

[4] Global Electric Car Count Climbs from 5.6 to 7.9 Million [Online]. Available at: <https://energie.themendesk.net/global-electric-car-count-climbs-from-5-6-to-7-9-million/> Accessed on February 1, 2021.

[5] Electric cars market statistics [Online]. Available at: <https://www.vpsolar.com/en/electric-cars-market-statistics/> Accessed on February 1, 2021.

[6] How Electric Vehicles Interact With a Smart Power Electricity Network – Vehicle to [Online]. Available at: <https://www.noodoe.com/how-electric-vehicles-interact-with-a-smart-power-electricity-network-vehicle-to-grid/> Accessed on February 3, 2021.

Protection system for microgrid and smart grid systems

[1]*Róbert Štefko (1st year)*
Supervisor: [2]*Michal Kolcun*

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

[1]robert.stefko@tuke.sk, [2]michal.kolcun@tuke.sk

***Abstract*— In recent years, the issue of the smart grid has become more significant because of the shortening of the time to decarbonize power engineering sources by 2050. With this fact, current challenges are emerging in the field of active protection systems and fault localization. The excellent aim is to amply provide unique designs using already deployed equipment in operation in protection system and fault localization or to efficiently develop intelligent facilities that will be made for this modern concept of smart grids.**

***Keywords* — microgrid, smart grid, renewable energy sources, fault localization, protection system.**

I. Introduction

In 2011, the European Commission presented a plan to reduce CO_2 emissions from power engineering by more than 80% by 2050. Therefore, increasing the efficiency of power sources and renewable energy sources will be crucial. The plan aims to significantly reduce carbon emissions by 2050 and ensure a stable energy supply. For this reason, changes will be necessary, which brings new challenges in the construction of a smart grid and microgrid.

Smart Grid is a system that makes certain the transmission of electricity from the producer to the consumer and at the same time controls and monitors current flows to achieve the highest energy efficiency and ensure a continuous supply of electricity. Smart Grid analyzes information accessed from local microgrid control equipment and accordingly optimizes power generation, reduces costs, and increases the reliability and efficiency of power organization.

Microgrid is a system designed to provide local control and monitoring. In the event of faults, switch to island mode and selectively rectify the fault location. For this reason, it is necessary to pay attention to the design of a suitable energy mix of power sources in each area to ensure the supply of electricity even in a purely island mode [1].

These changes equally affect the protection system and the location of faults. We currently have several devices for fault localization currently used in distribution networks. As the development of microgrid systems together with smart grids is gradually advancing, it is necessary to pay attention to these devices for locating faults together with the protection system and to propose new solutions that can be applied in these smart grid systems. This article describes some of these fault location devices in more detail.

II. Smart Grid and Microgrid

Key benefits of the smart grid are uninterrupted power supply for all consumers, reduced transmission and distribution loss, high penetration of renewable energy sources, cyber secured electrical network, large-scale power storage, flexibility to consumers to interact with energy market, market-based electricity pricing and demand side management [1]–[2].

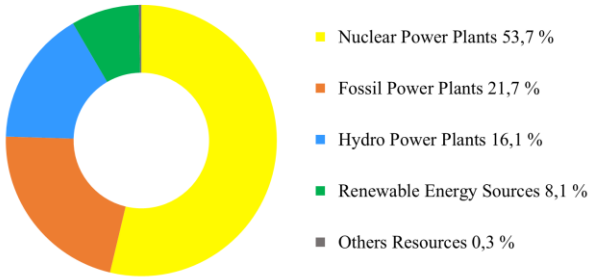


Fig. 1. Share of sources in electricity production in SR in 2019 [9].

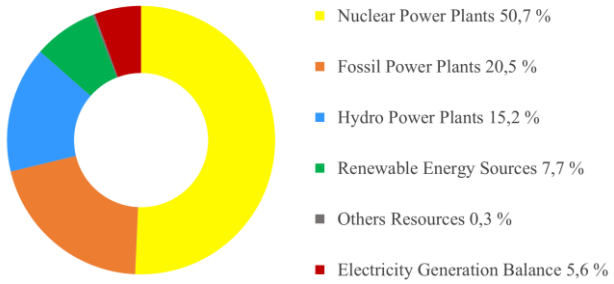


Fig. 2. Share of sources in electricity consumption in SR in 2019 [9].

When designing the smart grid, we must choose the appropriate size of local microgrids systems and not forget to select the right energy mix of sources that will power these local systems and therefore ensure a continuous supply of electricity due to the possibility of switching to island mode [4]–[5]. *Fig. 1* shows the share of all sources in production in the Slovak Republic in 2019, while the share of energy produced from renewable energy sources (RES) is 24.2%. Fig. 2 shows the share of all sources in total consumption, while

the share of RES is 22.9% and this figure could be increased by at least 5.6% of overall consumption in the Slovak Republic to achieve a balance between production and consumption [9]. Transmission system has constantly provided a balancing and management role in the electricity supply chain, while distribution networks have been designed to be passive (adapted and forgotten) in operation. The role of distribution companies is to take over the same services offered by the transmission system, such as energy flow control, contingency analysis and balancing in distribution networks. Distribution networks will need to be able to respond or adapt in real time to the complex interactions of all these challenges and to provide better information to various actors to enable different services to be traded in real time [6].

Key elements and priority components [6]:

• Simple, robust, secure, and flexible communication infrastructure to allow monitoring, management, control, and dispatching operations at all levels down to the distribution and customers.

• Common information and data models for all information building blocks, to ensure consistent database management, need to be defined at all levels of the power system and electricity supply chain.

• Well functioning Information and Communication Technology solutions are essential for maintaining the security of supply and for the efficient interaction of the market players.

• A truly competitive situation for all kinds of products relying on multi-vendor strategies can only be achieved with well-defined and standardized Information and Communication Technology solutions.

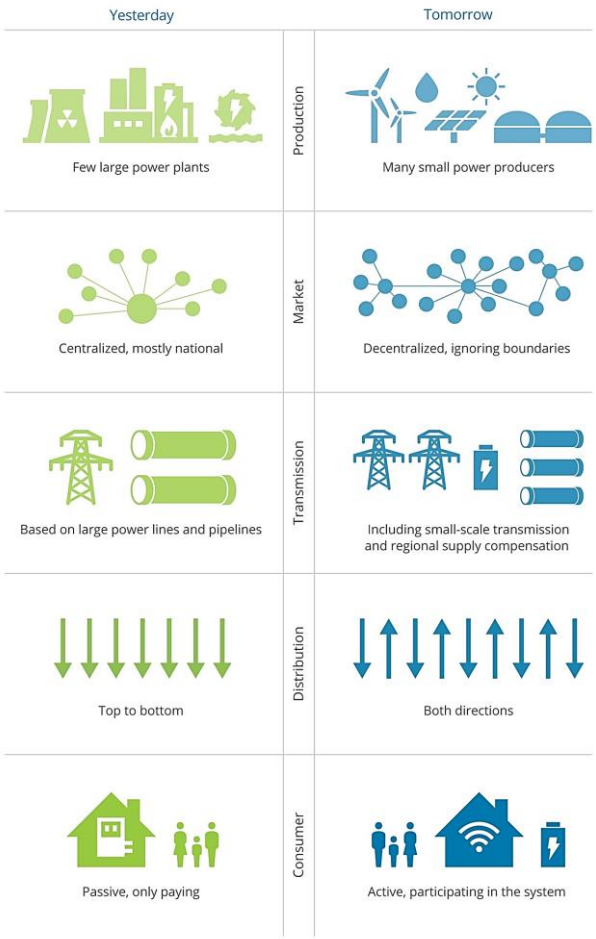


Fig. 3. Comparison of the current power engineering system on the left side with the smart grid on the right side [11].

Fig. 3 shows a comparison of the power engineering organization currently used with smart grids, with significant differences when looking at the *Fig. 3*. The difference in production is the transition from large power plants to many small power plants, which changes centralized production to decentralize, which is a significant change in smart grid systems. Another significant difference is distribution, where the transmission structure is lightened, and a distribution system with compensation is used.

While the direction of energy flow has a major impact on these facts. In the current system, we use only one direction, from production to consumption, but in smart grid, this fact changes into a two-way flow of energy, with the consumers themselves also participating in the production[2]–[10].

Among various grid-connected devices, smart meters are considered one of the most successful applications of the Internet of things (IoT) technology. In 2019, 14% of all the global meters were intelligent devices [11].

III. Renewable Energy Sources

The fluctuating and unpredictable nature of RES as photovoltaic station and the windmills require complex technologies to integrate to an existing grid. The harmonics developed from the complex control electronics circuits used for integration also make many troubles in the power system. Another disadvantage of renewable energy production is the need for a large area and a small short-circuit current contribution. The reliability of protection circuits to isolate it from the existing grid whenever required is a great challenge in the integration of RES [6]–[7]. The evolution in the power electronics to handle high power by fast semiconductor switching using IGBTs plays a major role in integration of renewable energy sources. These power electronics circuits may cause harmonics in the power system. The FACTs devices like static synchronous compensator (STATCOM), static series synchronous compensator (SSSC) and unified power flow controller (UPFC) will help to improve the stability and reliability of the grid which has large amount of renewable energy penetration. The filter circuits provided with these systems will mitigate the harmonics injected into the grid due to power electronics circuits. Realtime computer controllers which can implement advanced and complex algorithms also help to extract maximum power from renewable energy sources and reduce fluctuation in the renewable energy generation. It also protects the storage devices like battery from over charging. This will enhance the life of battery. It is more economical and grids friendlies if the fluctuated power generated by the renewable energy sources is utilized for irrigation purposes [8]. The fluctuation in renewable energy generation can also be reduced by distributed renewable energy generation to a larger geographical area in small power plants instead of the large power plant concentrating in one area. The output of solar photovoltaic may change 70% within five to ten minutes due to the local phenomenon like cloud passing. The vast land area requirement and the fluctuation of renewable energy can be minimized by distributed renewable energy generation [3]–[7]. Assigning a place to operate one of the RES technologies involves on many elements, that make integration under challenge. To begin with, some RES is unavailable in each region. Second, distance between RES and the grid remain a significant aspect in term of cost and efficiency. Moreover, RES depends on weather, climate, and geographical location. For example, solar energy in state of Arizona or Texas is more

efficient than in Michigan due to the significant percent of sun radiation time per day in Arizona and Texas. As a result, the integration will be more flexible in these states than Michigan [12].

IV. FAULT LOCALIZATION IN MICROGRID

The provision of power supply is equally affected by disturbances, the poor localization and elimination of which can cause cascade spread and, in the worst case, the disintegration of the electricity system and following blackout. In the possible future, the upcoming smart grid systems and the decentralization of RES should progressively eliminate these acute problems.

A. Use of a magnetic field to localization the fault area.

The measured magnetic field (B-field) is applied to an adaptive sensor. Slow variations in the load current will not affect the detector. A fault current will cause a rapid increase in the measured magnetic field which the sensor will respond to [13].

The detector will require two conditions are satisfied [13][14][15]:

1. The relative increase magnetic field percentage exceeds a configurable level.
2. The absolute increase magnetic induction (μT) is greater than a pre-set value.

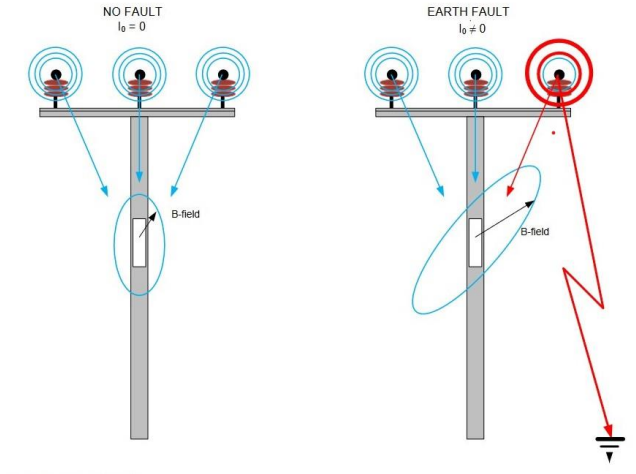


Fig. 4. Comparison of the change of the magnetic field in normal operation (left side) and in the fault state (right side) [13].

It is important to notice that the indicator measure a B-field which is a product of the current flowing in the feeder [13].

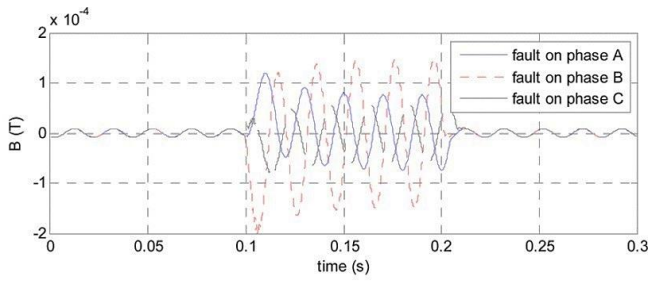


Fig. 5. Measured magnetic fields along different axes under single-phase short-circuit conditions. The figure shows the magnetic-field waveform along the x axis [14].

B. Embedded systems.

In recent years, the number of articles for the integration of embedded systems into distribution networks has increased significantly, which has a positive effect on progress in smart grids, while the issue of communication is still being addressed. Currently, the main wireless communication modes configured for the fault indicator are General Packet Radio Service (GPRS) and 3G/4G. However, there is limited or no communication coverage in remote mountainous or forest areas. Hence, it is impractical to deploy GPRS-based or 3G/4G-based fault indicators in the power distribution network in the abovementioned areas. It is of paramount importance to deal with this problem [16]–[18].

When LoRa and NB-IoT are combined and applied in the power distribution network, it can also be realized to achieve long-distance data transmission and low power consumption in fault localization [19]–[20].

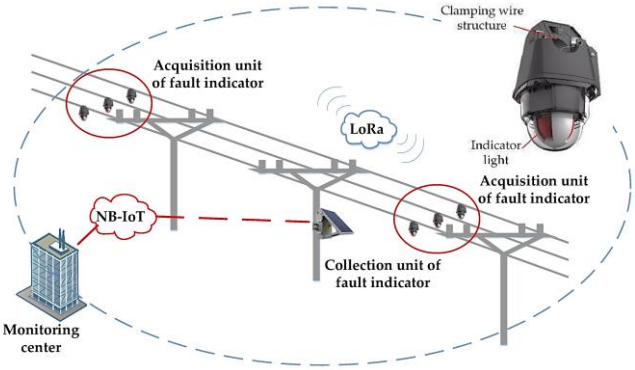


Fig. 6. The working principle of fault indicator based on LoRa and NB-IoT [16].

The operating principle of the intelligent fault indicator based on LoRa and NB-IoT is shown in *Fig. 6*. The acquisition unit of fault indicator samples current waveform of the line continuously at a frequency of 4 kHz. When a fault occurs on the line, the short-circuit and grounding fault measuring a module of the acquisition unit measures the fault signal on time and sends an interrupt signal to the single chip microcomputer, the single chip microcomputer responds to the interrupt, calibrates the fault occurrence time, and makes the acquisition unit store the data of 10 cycles before and after the fault occurrence time. Based on this data, the fault location can be judged whether it is upstream or downstream of the fault indicator. At that point, the acquisition unit sends the judgment information to the nearby collection unit by LoRa communication technology. The collection unit needs to summarize the judgment information sent by the acquisition unit to realize the fault location. The collection unit has LoRa and NB-IoT communication functions and can send the fault information of the line to the monitoring center by NB-IoT communication technology. Therefore, the staff can know the fault information of the entire distribution network in the monitoring center [16].

C. Detection, and fault localization using Petri Nets.

Petri nets were first invented by Carl Adam Petri in 1939. The invention was first used to get a good understanding of chemical reactions. Later the approach proved to be a good method in finding communication errors to a great extent. The same method has been proposed for finding fault location in power plants which is not being extended to a complete distribution network [21]. The algorithm has been tested creating three simultaneous faults in the network topology in the zones 2, 4 and 7. The relevant petri net token distribution are shown in *Fig. 7* [17].

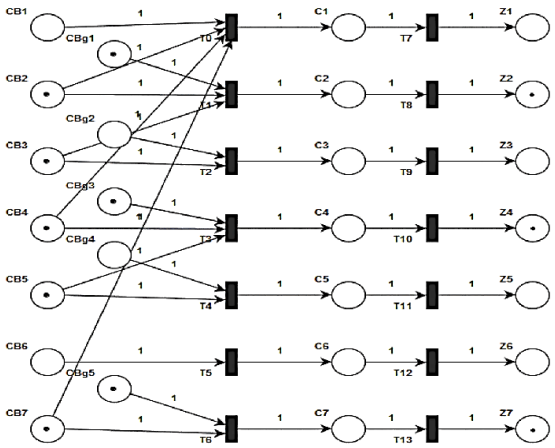


Fig. 7. Petri net model with token [17].

V. Conclusion

A major effort will be required for the successful rebuild of the current power structure to a smart grid. This article points out promising possibilities for use in protection and fault localization systems, which can help in the rebuild of the power system to a smart grid. In recent years, microgrid systems related to the smart grid have also gained attention. Therefore, it is necessary to draw thinking to research dealing with the selection of a suitable power mix of sources. The research shows that it is necessary to constantly pay attention to the development of communication of devices and smart grids systems. To successfully rebuild the current power system to a smart grid, it will be necessary to find answers to the following questions:

1. What is the appropriate source of energy mix for Slovakia according to geographical location?
2. What is the appropriate way of data processing and communication for a smart grid?
3. Is it necessary to develop new devices for protection and fault localization systems or is it enough to modify the current system?

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576. and Grant FEI no. FEI-2020-62 and by Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic (KEGA) under the project No. 008TUKE-4/2019.

References

[1] C. H. Ng, T. Logenthiran and W. L. Woo, "*Intelligent distributed smart grid network — Reconfiguration,*" 2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA), Bangkok, 2015, pp. 1-6, doi: 10.1109/ISGT-Asia.2015.7387125.

[2] A. A. Khadar, T. M. Hayath and M. S. Nagraj, "*Design and implementation of ZigBee based smart grid system for power management,*" 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon), Bangalore, 2017, pp. 1339-1343, doi: 10.1109/SmartTechCon.2017.8358584.

[3] C. P. Vineetha and C. A. Babu, "*Smart grid challenges, issues and solutions,*" 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG), Taipei, 2014, pp. 1-4, doi: 10.1109/IGBSG.2014.6835208.

[4] Xiaoling Jin, Yibin Zhang and Xue Wang, "*Strategy and coordinated development of strong and smart grid,*" IEEE PES Innovative Smart Grid Technologies, Tianjin, 2012, pp. 1-4, doi: 10.1109/ISGT-Asia.2012.6303208.

[5] L. Che, M. E. Khodayar and M. Shahidehpour, "*Adaptive Protection System for Microgrids: Protection practices of a functional microgrid system.,*" in IEEE Electrification Magazine, vol. 2, no. 1, pp. 66-80, March 2014, doi: 10.1109/MELE.2013.2297031.

[6] T. Vijayapriya and D. Kothari, "*Smart Grid: An Overview,*" Smart Grid and Renewable Energy, Vol. 2 No. 4, 2011, pp. 305-311. doi: 10.4236/sgre.2011.24035.

[7] A. S. Anees, "*Grid integration of renewable energy sources: Challenges, issues and possible solutions,*" 2012 IEEE 5th India International Conference on Power Electronics (IICPE), Delhi, 2012, pp. 1-6, doi: 10.1109/IICPE.2012.6450514.

[8] F. Liu, Q. Tang, T. Li, W. Xu, J. Gou and K. Li, "*Comparative Review of Flexible Alternative Current Transmission System Devices in the Smart Grid,*" 2019 IEEE 3rd International Electrical and Energy Conference (CIEEC), Beijing, China, 2019, pp. 194-198, doi: 10.1109/CIEEC47146.2019.CIEEC-2019105.

[9] SEPS, *yearbook 2019* [Online], [cit. 28.1.2021], Available on the Internet: <https://www.sepsas.sk/Dokumenty/RocenkySed/ROCENKA_SED_2019.pdf>.

[10] I. COLAK, R. BAYINDIR and S. SAGIROGLU, "*The Effects of the Smart Grid System on the National Grids,*" 2020 8th International Conference on Smart Grid (icSmartGrid), Paris, France, 2020, pp. 122-126, doi: 10.1109/icSmartGrid49881.2020.9144891.

[11] Infopulse, "*The Role of Smart Grid, IoT, and Big Data in Renewable Energy,*", article, 2020 [Online], [cit. 29.1.2021], Available on the Internet: <https://www.infopulse.com/blog/role-smart-grid-iot-big-data-renewables/>.

[12] A.K. Alsaif, "*Challenges and Benefits of Integrating the Renewable Energy Technologies into the AC Power System Grid*", American Journal of Engineering Research (AJER), vol. 6, no. 4, pp. 95-100, 2017.

[13] Nortroll AS, "*LineTroll R400D,*" user manual, 2013 [Online], [cit. 29.1.2021], Available on the Internet: <http://www.mr-bh.com/wp-content/uploads/2019/08/UG-LineTroll-R400D.pdf>.

[14] Q. Huang, W. Zhen, and P. W. T. Pong, "*A Novel Approach for Fault Location of Overhead Transmission Line with Noncontact Magnetic-Field Measurement,*" in IEEE Transactions on Power Delivery, vol. 27, no. 3, pp. 1186-1195, July 2012, doi: 10.1109/TPWRD.2012.2190427.

[15] M. Kazim, A. H. Khawaja, U. Zabit, and Q. Huang, "*Fault Detection and Localization for Overhead 11-kV Distribution Lines with Magnetic Measurements,*" in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 5, pp. 2028-2038, May 2020, doi: 10.1109/TIM.2019.2920184.

[16] Li, J.; Tang, J.; Wang, X.; Xiong, B.; Zhan, S.; Zhao, Z.; Hou, H.; Qi, W.; Li, Z. "*Optimal Placement of IoT-Based Fault Indicator to Shorten Outage Time in Integrated Cyber-Physical Medium-Voltage Distribution Network,*" Energies 13, 2020, no. 18:4928. doi: 10.3390/en13184928.

[17] R. Sreerama and K. S. Swarup, "*Detection, localization and fault diagnosis using PetriNets for smart power distribution grids,*" 2017 7th International Conference on Power Systems (ICPS), Pune, 2017, pp. 596-600, doi: 10.1109/ICPES.2017.8387363.

[18] Siemens, "*Fault Localization in the cloud – SICAM Localizer The watchful guardian for your overhead lines,*" article, [Online], [cit. 29.1.2021], Available on the Internet: <https://new.siemens.com/global/en/products/energy/energy-automation-and-smart-grid/distribution-automation/fault-localization.html>.

[19] B. Li, J. Wei, Y. Liang and B. Chen, "*Optimal Placement of Fault Indicator and Sectionalizing Switch in Distribution Networks,*" in IEEE Access, vol. 8, pp. 17619-17631, 2020, doi: 10.1109/ACCESS.2020.2968092.

[20] V. L. Vigni, A. Di Stefano, R. Candela and E. R. Sanseverino, "*A two-end traveling wave fault location system for MV cables based on LoRa technology,*" 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Milan, 2017, pp. 1-6, doi: 10.1109/EEEIC.2017.7977493.

[21] Z. Wang, F. Gao, X. Lv, W. Yin and J. Dong, "*A knowledge-based fault diagnosis platform in smart grid: A conceptual design,*" 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, 2014, pp. 1-5, doi: 10.1109/PESGM.2014.6938834.

Transient-State Temperature Calculation of Overhead Power Line Conductors

Anastázia MARGITOVÁ (4th year)
Supervisor: Michal KOLCUN

Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
anastazia.margitova@tuke.sk, michal.kolcun@tuke.sk

***Abstract*—This article deals with the temperature time course calculation of ACSR conductors according to the methodology stated in CIGRE Technical Brochure 601. Calculations of the conductor temperature time course were performed by three different calculation methods and the results were compared with the measurement under laboratory conditions These laboratory measurements were investigated by application of different current step change with and without the wind blowing.**

***Keywords*—CIGRE Technical Brochure 601, temperature time course calculation, ACSR conductor.**

I. Introduction

One of the most important factors influencing the operation of overhead power lines is the conductor temperature [1]. The temperature of the conductor is affected by several factors, which can be divided into two groups [2], [3]:

- characteristics of overhead power lines (current flowing through conductors, type, geographical location, sag, tension of conductors, ground clearance, etc.),
- climatic conditions in which the power lines are located (wind speed and direction, ambient temperature, global solar radiation, precipitation, etc.).

Dynamic thermal rating of overhead power lines provides a real conductor ampacity based on operating parameters (climatic conditions and current flowing through the conductor) in real time. If the conductor has a certain heat capacity, it can be used for short-term current overload without exceeding its maximum allowable temperature. This short-term current overload represents a transient state and is related to a limited overload time. This method of calculation is focused on the change of the conductor temperature with respect to time, it takes into account changes in operating parameters over time [4].

II. Mathematical Model for Calculation of Conductor Temperature in the Transient State

According to CIGRE Technical Brochure 601 (TB 601), the general heat equation of the conductor in the case of the transient-state can be expressed by Eq. (1) [5]:

$$m \cdot c \cdot \mathrm{d}T_\mathrm{s}/\mathrm{d}t = P_\mathrm{J} + P_\mathrm{s} - P_\mathrm{c} - P_\mathrm{r}\,, \qquad (1)$$

where

m is the conductor mass per unit length (kg/m),
c is the conductor specific heat capacity (J/(kg·K)),
$\mathrm{d}T_\mathrm{s}/\mathrm{d}t$ is the conductor temperature time change (°C/s),
P_J is the Joule heating (W/m),
P_s is the solar heating (W/m),
P_c is the convective cooling (W/m),
P_r is the radiative cooling (W/m).

Analytical Calculation of Conductor Temperature

The easiest way to model changes in operating parameters is a step change. The conductor temperature does not change immediately after a step change in operating parameters, but it takes some time. In the case of an analytical solution, this change is mathematically represented by an increasing or decreasing exponential function expressed by Eq. (2) [4], [5]:

$$T_\mathrm{s} \cong T_\mathrm{sm} - \left(T_\mathrm{sm} - T_\mathrm{s1}\right) \cdot \exp\left(-t/\tau\right), \qquad (2)$$

where

T_s1 is the conductor temperature at time $t = 0$ (°C),
T_s is the conductor temperature at time t (°C),
T_sm is the conductor temperature at time $t = \infty$ (°C),
τ is the global thermal time constant (s),
t is time (s).

The global thermal time constant includes the effects of changes in all operating parameters, i.e. in climatic conditions, but also in the current flowing through the conductor and it can be determined from Eq. (3) [4], [5]:

$$\tau \cong m \cdot c \cdot \left(T_\mathrm{sm} - T_\mathrm{s1}\right) \Big/ \left(\Delta P_\mathrm{J} + \Delta P_\mathrm{s} - \Delta P_\mathrm{c} - \Delta P_\mathrm{r}\right), \qquad (3)$$

where

ΔP_J is the change in Joule heating after a step change in current (W/m),
ΔP_s is the change in solar heating after a step change in global solar radiation (W/m),
ΔP_c is the change in convective cooling after a step change in ambient temperature, wind speed or wind direction (W/m),
ΔP_r is the change in radiative cooling after a step change in ambient temperature (W/m).

Numerical Calculation of Conductor Temperature

The numerical method of conductor temperature calculation represented by Eq. (4), also called temperature tracking, is based on the numerical expression of Eq. (1) [4], [5]:

$$T_\mathrm{s}^{(k+1)} = T_\mathrm{s}^{(k)} + \left(\left(P_\mathrm{J}^{(k)} + P_\mathrm{s}^{(k)} - P_\mathrm{c}^{(k)} - P_\mathrm{r}^{(k)}\right) \Big/ \left(m \cdot c\right)^{(k)}\right) \cdot \Delta t\,, \qquad (4)$$

where

k is the number of iterations (–),
Δt is the time step (s).

Eqs. (1) to (4) represent basic equations that describe the thermal behavior of a conductor in the transient state The equations for the calculation of the quantities given in Eqs. (1) to (4), as well as a detailed explanation of the analytical and numerical method for calculating the conductor temperature time course are given in [4]-[6].

III. Conductor Temperature Calculation Based on Real Measurement under Laboratory Conditions

This section deals with the calculation of conductor temperature time course according to TB 601 based on real measurement in laboratory conditions. As presented in [6] by M. Kanálik, A. Margitová, Ľ. Beňa, several measurements were performed for two types of ACSR conductors. During these measurements, step changes in the current flowing through the conductor were made, as well as a step change in the wind speed, which led to a change in the conductor temperature. In [6], the authors deal with the calculation of conductor temperature only in a steady state, the conductor temperature was calculated only after its stabilization. Then, steady-state temperature dependencies on the current flowing through the analyzed conductors were created.

The paper [4] expands the issue examined in [6] by the calculation of the conductor temperature in the transient state, i.e. it takes into account changes in the conductor temperature over time. The conductor temperature change over time was calculated by three methods: analytical solution, simplified temperature tracking, and detailed temperature tracking. A more detailed description of the given measurements and calculation methods is given in [4], [6].

Measured and calculated steady-state temperature of analyzed conductors as well as the considered operating conditions for two selected measurements are shown in TABLE I. Measurements were divided into two time intervals (I. and II.). At the beginning of the first interval, a step change in current at a wind speed of 0 m/s was considered. At the beginning of the second interval, a step change in wind speed to 2 m/s were considered.

TABLE I
Measured and Calculated Steady-State Temperature, Operating Parameters for Two Selected Measurements

Type of ACSR conductor		Time (min)	Wind speed (m/s)	Wind direction (°)	Global solar radiation (W/m^2)	Measured average ambient temperature (°C)	Measured average current (A)	Measured steady-state conductor temperature (°C)	Calculated steady-state conductor temperature (°C)
429-AL1/52-ST1A		$t = 0$	0	90	0	23	0	22.2	23
	I.	$0 < t \leq 290.1$	0			23	1001.3	108.7	110
	II.	$t \geq 290.1$	2			24.5	1013.5	49.7	52
352-AL1/59-ST1A		$t = 0$	0	90	0	22.3	0	21.7	22.3
	I.	$0 < t \leq 297.7$	0			22.3	698.3	78	77.7
	II.	$t \geq 297.7$	2			23.5	703.3	40.5	39.7

Only the step changes in the operating parameters listed in TABLE I were considered in the case of the analytical solution and simplified temperature tracking. However, the ambient temperature and current were not constant during the measurements for analyzed time intervals. Considering the average values of measured ambient temperature and current for analyzed time intervals therefore caused some inaccuracy in the calculations for these two methods. A time step of 1 s was considered for simplified and detailed temperature tracking. In the case of detailed temperature tracking, the input values of measured ambient temperature and current were updated for each second.

Conductor current, conductor temperature and ambient temperature time variations during two selected measurements are shown in Fig. 1 and Fig. 2. The calculated conductor temperature by analytical solution, simplified temperature tracking, and detailed temperature tracking is also shown in Fig. 1 (for conductor 429-AL1/52-ST1A) and Fig. 2 (for conductor 352-AL1/59-ST1A). The relative deviation of calculated conductor temperature (by three above-mentioned methods) from measured data for the presented measurements (Fig. 1 and Fig. 2) is shown in Fig. 3 and Fig. 4. The individual figures show a good correlation between the calculated and measured values of the conductor temperature, which indicates that the calculation results performed by all three methods are close to real measurements.

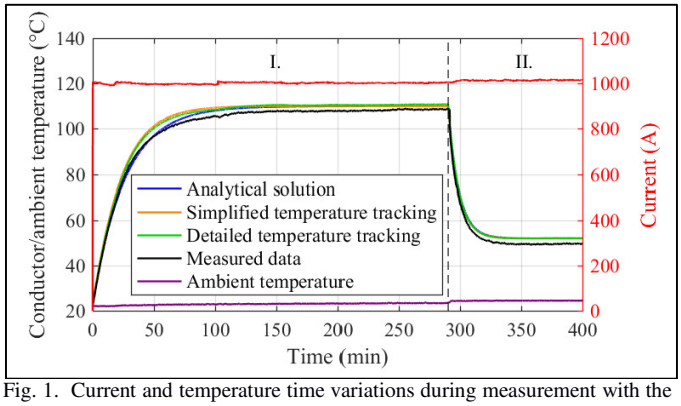


Fig. 1. Current and temperature time variations during measurement with the conductor 429-AL1/52-ST1A [4].

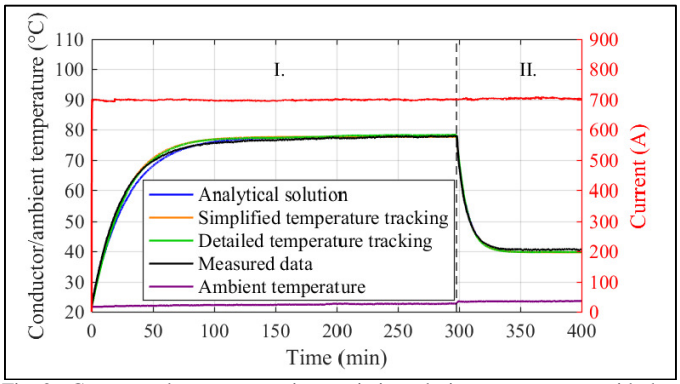


Fig. 2. Current and temperature time variations during measurement with the conductor 352-AL1/59-ST1A [4].

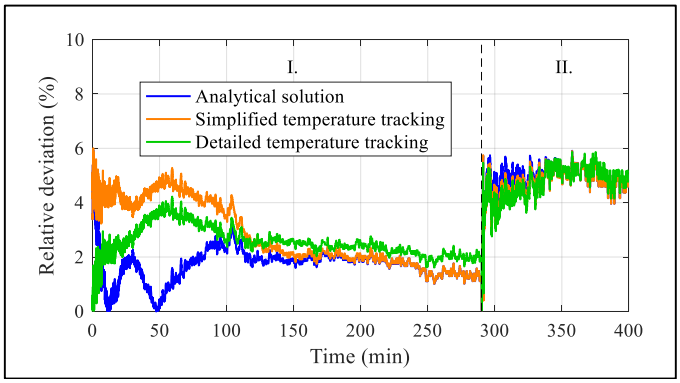


Fig. 3. Relative deviation from measured data for conductor 429-AL1/52-ST1A [4].

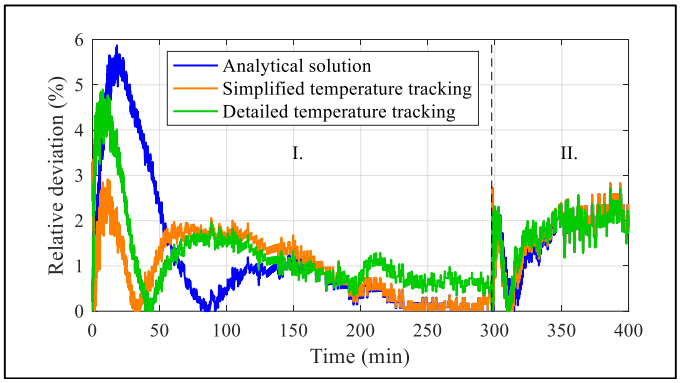


Fig. 3. Relative deviation from measured data for conductor 352-AL1/59-ST1A [4].

IV. Conclusion

In this article, the calculation of the conductor temperature time course can be realized in three different ways: analytical method, simplified/detailed temperature tracking. For all these methods, the key values are the operating parameters before and after the step change. An analytical method is based on steady-state values of the conductor temperature before and after a step change of the given operating parameters. The advantage of temperature tracking is that this method makes it possible to estimate the temperature of the conductor accurately quite enough even in the case of changes in operating parameters during the transient, i.e. during the state when the conductor is not in thermal equilibrium. Further research will be focused on the use of the temperature tracking method for a real overhead power line ampacity computation in the case of a long time step interval of input data measurement, which usually is about 10 or 15 minutes.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576.

References

[1] S. Beryozkina, A. Sauhats, and E. Vanzovichs, “Climate conditions impact on the permissible load current of the transmission line,” in *2011 IEEE Trondheim PowerTech*, Trondheim, Norway, 2011, pp. 1–6.

[2] S. Abbott, S. Abdelkader, L. Bryans, and D. Flynn, “Experimental validation and comparison of IEEE and CIGRE dynamic line models,” In: *45th International Universities Power Engineering Conference (UPEC2010)*, Cardiff, Wales, UK, 2010, pp. 1–5.

[3] S. Karimi, P. Musilek, and A. M. Knight, “Dynamic thermal rating of transmission lines: a review,” in *Renewable and Sustainable Energy Reviews*, vol. 91, 2018, pp. 600–612.

[4] M. Kanálik and A. Margitová, “Transient-state temperature calculation of overhead power line conductors based on CIGRE Technical Brochure 601,” In *Arabian Journal for Science and Engineering*, submitted for publication.

[5] CIGRE, Working Group B2.43, Guide for thermal rating calculation of overhead lines, Technical Brochure 601, 2014.

[6] M. Kanálik, A. Margitová, and Ľ. Beňa, “Temperature calculation of overhead power line conductors based on CIGRE Technical Brochure 601 in Slovakia,” In *Electrical Engineering*, vol. 101, 2019, pp. 921–933.

Monitoring of power system using WAMS technology – during disturbances

[1]*Vladimír KOHAN (2nd year)*
Supervisor: [2]Michal KOLCUN

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

[1]vladimir.kohan@tuke.sk, [2]michal.kolcun@tuke.sk

***Abstract*— The article is focused on the display of the courses of individual monitored quantities using WAMS systems in the Slovak electricity transmission system in the time span when there was a major outage of electricity production in continental Europe in neighboring Austria. After a more detailed description of the event from January 8, 2021, simulations from phasor units of PMU measurement from 400 kV power stations are plotted using the PhasorPoint program, which focus on how the controllers or dispatching reacted to ensure a steady state - a look at changing values of individual parameters. at a given point before failure to a steady state. The paper points out the functions and reactions of WAMS systems in the event of faults and simulations. It shows how the development of control shifts from the level of static stability to dynamic stability.**

***Keywords*— WAMS, PMU, power system, dynamic stability, synchrophasor.**

I. INTRODUCTION

On January 8, 2021 at 14:05 Central European Time, the synchronous area of continental Europe was divided into two separate areas due to the failure of several elements of the transmission network in a very short time. The route where the two areas were separated is as follows [12]:

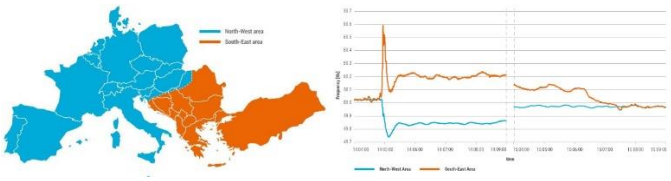


Fig. 1 Map of Continental Europe showing two separate areas during the systemic event and frequency in Continental Europe during the event on January 8, 2021

System division registered in the synchronous area of continental Europe - incident overview:

At approximately 14:05 Central European Time, the frequency in the northwestern part of continental Europe initially decreased to 49.74 Hz in about 15 seconds. Then the frequency reached a steady value of approximately 49.84 Hz. At the same time, the frequency in the southeast region initially increased to up to 50.6 Hz and then stabilized at a stable steady frequency between 50.2 Hz and 50.3 Hz. This can also be observed from the graph below [12].

II. EVENT DESCRIPTION

Due to the reduction in frequency in the north-west, contracted support services (discontinued) were disconnected in France and Italy, for a total of around 1.7 GW, to reduce frequency deviation. These services are disconnected by large customers who are contractually bound by the relevant transmission system operators (TSOs) if the frequency falls below a certain threshold. In addition, 420 MW of support capacity and 60 MW of UK support capacity were automatically activated from the Nordic synchronous areas. These countermeasures ensured that as early as 14:19 (CET) the frequency deviation was limited to a deviation of about 0.1 Hz in the north-western region from the nominal frequency of 50 Hz (see Fig. 1) [12].

Due to the high (excessive) frequency in the southeastern region, automatic and manual countermeasures were activated (i.e., reduced power supply to production units). Thus, at 15:05 (CET), the frequency deviation in the southeast region could be limited to +0.1 Hz from the nominal frequency of 50 Hz (see Fig. 1) [12]. The automatic response and the coordinated action taken by transmission system operators (TSOs) in Continental Europe have ensured that the situation is quickly restored to normal operation. Activities were coordinated between the TSOs in order to reconnect the two areas as quickly as possible. Therefore, contractual support (interruptible) services could reconnect at Italy at 14:47 (CET) and in France at 14:48 (CET). At 15:08 (CET), the two areas were again merged into one synchronous area in Continental Europe [12].

The area in the south-eastern region of the interconnected network was separated from the rest of Continental Europe during this period. A temporary decrease in frequency of approximately 250 mHz was noted. Coordinated action and immediate response by Continental European Transmission System Operators have ensured that system stability has not been affected in larger European countries [13].

Due to a sharp drop in frequency in the electricity transmission system, there was a threat of a massive blackout outage in Europe on Friday. With reference to the energy supplier Wien Energie and EVN, the APA informed about it. To stabilize the grid, Austrian power plants had to supply energy immediately. According to EVN, electricity from

nuclear and coal-fired power plants in neighboring countries has also helped Austria to cope.

III. Evaluation of the situation in the PhasorPoint program - using PMU units

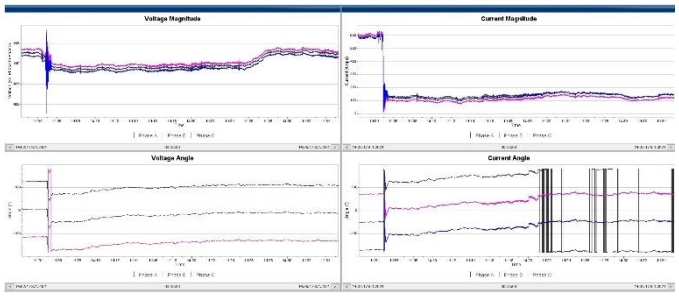


Fig. 2 Courses of combined voltages and voltage angles in the substation on the overhead line and curves of the magnitude of currents and current angles in the distribution station on the overhead line

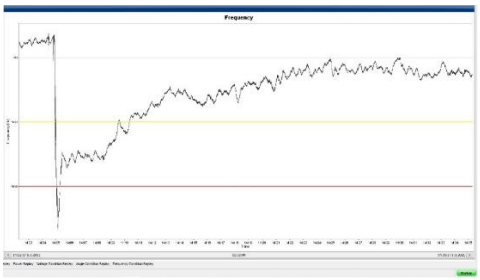


Fig. 3 Frequency course in the substation on the overhead line

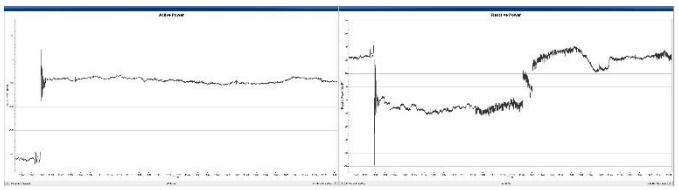


Fig. 4 The course of active power in the substation on the overhead line and course of reactive power in the substation on the overhead line

The graphs (Fig. 2-Fig. 4) show the courses of voltages, currents, frequency, active and reactive power in the power substation at the event of 08.01.2021. With the help of the PhasorPoint (GE) program and the suitable location of the measuring phasor units PMU, it is possible to see continuously (according to the sampling frequency) the courses of all observed quantities. Such advantages are also provided by WAMS systems, which can eliminate and reduce the risk of e.g. BlackOut with early warning (alarm) and earlier (more accurate) response of dispatchers thanks to a more thorough display (at a dynamic level) than previous SCADA / EMS interfaces [2], [5], [9], [10].

Conclusion

The research is focused on the appropriate use of PMU units in the power system in order to monitor and control the operation and modernize the dispatching - the transition from SCADA to the PMU system; i.e. from the area of visibility of static stability only (2-4 samples per second) to the area of visibility of dynamic phenomena (60+ samples per second). In this way, it is possible to control the operation more safely, which will then be reflected in a high-quality and stable supply of electricity. Last but not least, the work deals with the initial analysis of the assessment of the current possibilities of using the WAMS system for dispatch control in PhasorPoint, detection of island operation, control of oscillation damping, detection of generator oscillation, evaluation and prediction (estimation) of system status and alarm and warning settings [1], [3], [4], [8].

References

[1] LI, Y. – YANG, D. – LIU, F. – CAO, Y. – REHTANZ, CH. 2016. Interconnected Power Systems; Wide-Area Dynamic Monitoring and Control Applications, Berlin Heidelberg: Springer-Verlag, 2016. 98-112 p. ISBN 978-3-662-48625-6.

[2] MONTI, A. – MUSCAS, C. – PONCI, F. 2016. Phasor Measurement Units and Wide Area Monitoring Systems; From the Sensors to the System, London: Academic Press is an imprint of Elsevier, 2016. 2-111 p. ISBN: 978-0-12-804569-5.

[3] MA, J. 2018. Power system; Wide-Area Stability Analysis and Control, Beijing-China: Wiley-Science Press Beijing, 2018. 23-154 p. ISBN 9781119304876.

[4] PHADKE, A. G. – THORP, J. S. 2017. Synchronized Phasor Measurements and Their Applications, Second Edition, USA: Springer International Publishing, 2017. 14-243 p. ISBN 978-3-319-50582-4.

[5] MESSINA, A. R. 2015. Wide-Area Monitoring of Interconnected Power Systems, London: The Institution of Engineering and Technology, United Kingdom, 2015. 2-98 p. ISBN 978-1-84919-853-0.

[6] VACCARO, A. – ZOBAA, A. F. 2016. Wide Area Monitoring, Protection and Control Systems; The enabler for Smarter Grids, London: The Institution of Engineering and Technology, United Kingdom, 2016. 3-86 p. ISBN 978-1-84919-830-1.

[7] BEVRANI, H. – WATANABE, M. – MITANI, Y. 2014. Power System Monitoring And Control, New Jersey: Published by John Wiley & Sons, Inc, Hoboken-IEEE Press, 2014. 31-185 p. ISBN 978-1-118-45069-7.

[8] SANTOS, L. F. – ANTONOVA, G. – LARSSON, M. – SERGIO, F. 2015. The Use of Synchrophasors for Wide Area Monitoring of Electrical Power Grids. [online]. Semantic Scholar: 1-15 p. [cit. 2020.09.21] Available on inernet: https://library.e.abb.com/public/9357d370a88948e3adcd1fdc1a216741/The%20Use%20of%20Synchrophasors%20for%20Wide%20Area%20Monitoring%20of%20Electrical%20Power%20Grids.pdf.

[9] BABNIK, T. – GÖRNER,K. – MAHKOVEC, B. 2014. Wide Area Monitorig System. [online]. Berlin Heidelberg: Springer-Verlag, 1-18 p. [cit. 2020.07.10] Available on inernet: https://www.kth.se/social/files/5829b140f276540cc86d5d1a/9%20WAMS.pdf.

[10] ABB. 2012. Wide Area Monitoring Systems; Portfolio, applications and experiences. [online]. ABB Group. 1-35 p. [cit. 2020.10.27] Available on inernet: https://library.e.abb.com/public/94fab39c67b4ac00c125784f00293520/1KHL501042%20PSGuard%20WAMS%20Overview%202012-04.pdf.

[11] ABB. 2015. Jornadas Técnicas Peru 2015 Protection, monitoring and control solutions using Wide Area Monitoring Systems. [online]. ABB Group. 1-29 p. [cit. 2020.11.07] Available on inernet: https://new.abb.com/docs/librariesprovider78/documentos-peru/presentaciones-primeras-jornadas-tecnicas-abb-peru/ps/wide-area-network-technology-and-synchrophasors.pdf?sfvrsn=2.

[12] ENTSO-E. 2021. System Separation in the Continental Europe Synchronous Area on 8 January 2021 – update. [online]. ENTSO-E. [cit. 2021.01.16] Available on inernet: https://www.entsoe.eu/news/2021/01/15/system-separation-in-the-continental-europe-synchronous-area-on-8-january-2021-update/

[13] ENTSO-E. 2021. System split registered in the synchronous area of Continental Europe – Incident now resolved. [online]. ENTSO-E [cit. 2021.01.16] Available on inernet: https://www.entsoe.eu/news/2021/01/08/system-split-registered-in-the-synchronous-area-of-continental-europe-incident-now-resolved/

[14] VANFRETTI, L – BAUDETTE, M – DOMÍNGUEZ-GARCÍA, J-L – ALMAS, M. S – WHITE, A – GJERDE, J. O. 2016. A Phasor Measurement Unit Based Fast Real-time Oscillation Detection Application for Monitoring Wind-farm-to-grid Sub–synchronous Dynamics, Stockholm: Electric Power Components and Systems, 2016. 123-134 p. ISSN: 1532-5016.

Synthetic End-user Trajectory Generation

[1]*Marek RUŽIČKA (2nd year),*
Supervisor: [2]*Juraj GAZDA*

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

[1]marek.ruzicka@tuke.sk, [2]juraj.gazda@tuke.sk

***Abstract*—The ongoing changes in RAN networks as 5G was deployed in 2020 create a need to dynamically respond to a fast-changing end-user behavior. This work proposes an approach to generate a huge synthetic dataset of end-user mobility that can be used for machine learning purposes.**

***Keywords*—5G, mobility, generation, dataset**

I. Introduction

According to Cisco annual internet report [1], between the years 2018-2023, we can expect 1.5-2 times growth in the number of mobile devices. There will be 25-50% of the devices connected to a RAN network. In central Europe, this means 1 billion devices connected to the RAN network, whereas in 2018 it was 0.6 billion. This massive growth in connections will lead to enormous traffic.

The 4G infrastructure is slowly replaced by the new 5G network. The 5G is relying highly on small cells [2]. The high-frequency signal used in 5G is more susceptible to Line-of-Sight (LOS) problems and can successfully travel way lower ratio. Considering that people are moving, a dynamic network architecture seems to be a good solution. As can be seen during the Covid-19 pandemic, human habits can change rapidly. As the Google mobility trends report shows, lock-downs and restrictions around the world have caused that mobility in the urban area dropped by 20-30%, while the mobility in residential locality increased by 15-20% and in parks, the mobility grew by 109% at its peak on August 12, 2020 [3]. Such events indicate that there can be a demand for a better connection in some situations that can not be covered by static infrastructure that is currently in use. The possible solution for such scenarios, where the end-users ground terminals (GT) move to the territory with bad signal coverage is to use a dynamic network of moving base stations - either on the ground or in form of unmanned aerial vehicles (UAVs), often referred to as drones. However, in such a scenario, there is a demand for efficient algorithms for moving the UAVs to the best possible location. This paper focuses on generating a synthetic dataset for machine-learning (ML) purposes with the aim to efficiently position the UAVs in the area.

II. Available datasets

In a 5G environment, obtaining a dataset is not a major problem, the positions are captured in real-time by the operator. However, such a dataset contains sensitive data and operators must follow local regulations. In common, the operators are not willing to share their data. A possible solution is to create a new dataset by getting volunteers to be tracked. Such data collection is however limited to the number of volunteers, their routes, and habits. Another option is to obtain a dataset from publicly available sources. Website data.world [4] provides 84 mobility datasets for various use cases. However, those datasets are very limited. Some of the datasets are related to traffic signalization, others are related to bike or sports activities, several address mobility in means of traveling, or simply are surveys on mobility trends. We found several datasets that partly fit our needs in means of data structure, but the density of the trajectories was low and there was no data at all for some parts of the selected area. The best-fitting dataset we were able to find was a taxi trajectories dataset from Beijing, China named T-Drive [5], [6], [7]. This dataset is however very inconsistent and barely reflects pedestrian motion. For this reason, generating a new dataset based on simulations reflecting real end-users habits comes as a good option.

III. End-user motion simulations

Trajectory can be formally defined as a sequences of GT positions $T = x_{t_1}, x_{t_2}, ..., x_{t_n} | x_i = (latitude_i, longitude_i)$, where n represents the number of time frames. To generate a trajectory dataset we need to create a simulator that can create realistic samples. Analysis of the GT motion is necessary. The average human speed when walking Δv is established to be around 5km/h, 6km/h is considered a fast walk, and speed above is usually running speed [8]. Speed under 5km/h is a slow walk. Keeping this in mind, we need to set a random speed to the simulated GTs with median value $\Delta v = 5km/h$, minimal value $v_{min} = 3km/h$ and maximal value $v_{max} = 9km/h$ for runners. We do not consider cars that are expected to be equipped with MFemtocells in the future [2]. MFemtocells are vehicle-mounted cells that are covering users inside the vehicle and connecting to a macrocell. Cyclists are not considered, the use of data transitions while riding can bring up legal issues and road traffic hazards. The second important parameter is the time between two measurements Δt. In the T-Drive dataset, the measurement time was very variable, from measures at each second to up to several minutes. Because of the slow speed of considered end-users, we can not expect the average distance Δd to exceed values more than 2,5m (1,4m on average) per second, which is not a reasonable value for the needs of ML purposes. Because the end-user can theoretically pass 150m in a minute, averaging at about 85 meters, this seems also as a high value. To select the best fitting Δt we also considered the practical use of the data. Because we need the data for machine learning

purposes and we track 24h periods of end-users behavior, we consider the reasonable count of measurements of end-users positions to be a power of two. Because a day has 1440 minutes, we need a higher number than this. The closest power of two to this value is 2048, which would mean that we need to save end-users' position each $\frac{1440}{2048} min \approx 42s$. In $t = 42s$ an average human moves $d = 60m$, which is a reasonable distance for RAN re-orchestration. The last important feature to be considered is the area to be simulated. Analyzing other works that are using end-user mobility data to predict future trajectories or UAV positions such as [9] or [10], the trajectories are located within a relatively small restricted area. In the same manner, we have chosen to cover an area of one square mile in Košice, Slovakia. The area in which the GTs move was restricted within latitude and longitude bounds as follows: $lat \in < 48.7171252, 48.7027684 >, lon \in < 21.2497423, 21.228062 >$. Because we model a theoretical end-user mobility pattern generator, we consider all users to stay within the given bounds. In real conditions, some users leave the area and others enter. The average mobility does not change if we do not consider this fact.

To simulate all the trajectories, the simulator is not only simulating one but n users. n is an input parameter of the simulator. Along with n, lat/lon bounds enter the simulator as well. The simulator then creates n GTs with the routine presented in algorithm 1. Each of the GTs gets one of the roles - scholar (pupils, students, etc.), working (people having casual working habits - create semi-iterative trajectories to/from work), and random. Random class is representing neither of the previous categories, random can stand for free time activities, seniors, people with no daily job, etc. Each GT in the simulator gets assigned random home coordinates from the restricted area. We used an OSMNX library [11] to select valid house/flat coordinates for this purpose. The same was used for selecting random coordinates of working buildings (business centers, restaurants, factories, etc.), and/or schools. Each GT gets a set of up to five random places where they can move with random probability representing hobby, free time activities, friend/family visits, etc. The probability for these actions is always lower than 15%. Random GTs have their work or school coordinates set, but they do not visit it periodically but on random occasions. periodic GTs (scholar/working) do attend their school or work in periodical time with a set duration of their visit to 8 hours, while in case of students this can vary by 25%. After leaving the school/work, simulated GT moves either home or to one of his favorite places where he spends a random period of up to 6 hours. The simulated moving from place A to place B is realized through the closest route calculated with the OSMNX library.

All GTs are moving independently, and each period Δt the information about their position is saved. We have been able to simulate thousands of GT trajectories with this approach and have created a robust dataset for possible ML usage.

Algorithm 1 Initializing the end users.

1: Require number of end-users n
2: Initialize end-users roles $R = \{scholar, working, random\}$
3: Require coordinate bounds $B = \{lat_{min}, lat_{max}, lon_{min}, lon_{max}\}$
4: Create empty users set $U = \emptyset$
5: **for** $n > 0, n = n - 1$ **do**
6: Create user u
7: Set a random role to user from R
8: Set random home coordinates within the bounds B for user u
9: Set random work/school coordinates within the bounds B for user u
10: **for** $it = random(1, 5), it > 0, it = it - 1$ **do**
11: Select random point r within the given bounds B and random probability $p < 0.15$
12: Assign the point r to set of users u favorite places $u_f = u_f + r$ with random probability $u_{f_{proba}} = u_{f_{proba}} + p$
13: Append the created user to the set of end-users $U = U \cup \{u\}$

IV. Conclusion and Discussion

The GT mobility patterns play a key role in the process of 5G network design. Massive increase of the RAN infrastructure complexity creates additional requirements on precise network planning and overall orchestration of the network as such. The possible solution to enhance the statistics feeding the network planning process is to generate massive dataset of the GT mobility patterns.

Many existing mobility datasets are not suitable for ML purposes of network orchestration. The GT mobility simulator was described in this paper and a huge dataset of simulated data was created.

This dataset can be applied to feed a neural network to generate new realistic trajectories in way lower time and time-slices of the generated data can be used to teach a neural network how to correctly orchestrate the RAN network for better throughput and lower latency.

Acknowledgment

This work was supported by the Slovak Research and Development Agency, project number APVV-18-0214 and by the Scientific Grant Agency of the Ministry of Education, science, research and sport of the Slovak Republic under the contract: 1/0268/19.

References

[1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.
[2] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE communications magazine*, vol. 52, no. 2, pp. 122–130, 2014.
[3] H. Ritchie, "Google mobility trends: How has the pandemic changed the movement of people around the world?" https://ourworldindata.org/covid-mobility-trends, accessed: 2021-01-19.
[4] data.world. (2021) Mobility datasets available on data.world. [Online]. Available: https://data.world/datasets/mobility
[5] Y. Zheng, "T-drive trajectory data sample," 8 2011, t-Drive sample dataset. [Online]. Available: https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/
[6] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 316–324.
[7] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, 2010, pp. 99–108.
[8] A. Thorstensson and H. Roberthson, "Adaptations to changing speed in human locomotion: speed of transition between walking and running," *Acta Physiologica Scandinavica*, vol. 131, no. 2, pp. 211–214, 1987.

[9] Y. Y. Munaye, H.-P. Lin, A. B. Adege, and G. B. Tarekegn, "Uav positioning for throughput maximization using deep learning approaches," *Sensors*, vol. 19, no. 12, p. 2775, 2019.

[10] H. A. Fayed and A. F. Atiya, "A mixed breadth-depth first strategy for the branch and bound tree of euclidean k-center problems," *Computational Optimization and Applications*, vol. 54, no. 3, pp. 675–703, 2013.

[11] G. Boeing, "Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, 2017.

Dynamic systems control of asynchronous motor drive

[1]*Marek FEDOR (1st year)*
Supervisor: [2]Daniela PERDUKOVÁ

[1,2] Department of Electrical Engineering and Mechatronics, Technical University of Kosice, Slovak Republic

[1]marek.fedor4@student.tuke.sk, [2]daniela.predukova@tuke.sk

***Abstract*— This work is focused on the design of a linear stable reference model control structure for continuous nonlinear systems, the stability conditions of which were determined in accordance with Lyapunov´s theory. This structure will be further applied in the control of angular speed of an asynchronous motor drive and compared with classic vector control. In terms of control, an induction motor drive represents a highly nonlinear higher order system. Comparison of scalar control method with classic vector control will be on the basis of stability, invariance and robustness.**

***Keywords*—Dynamics of Speed Control, Reference Model Structure, Induction Motor Drive, Nonlinear System.**

I. INTRODUCTION

Induction or asynchronous motors are a type of electrical drives powered by an alternating current and converting electrical energy – from either stored power or direct electrical connection – into mechanical energy through production of rotational force. They are used widely because of their robustness, reliability and low maintenance. However, controlling an induction motor is rather difficult due to variable frequency, complex dynamic and parameter variations, etc. [1] – [3]. In terms of control, the induction motor presents a typical example of a nonlinear system, as it is a significantly oscillating system incorporating also positive feedback. Besides this it is very often influenced by unknown external disturbances and its parameters tend to change depending on the temperature, frequency and other operational conditions, all of which significantly influence the accuracy of its speed control.

The scalar speed control method has a simple control structure [4], [5] and is implemented easily in general-purpose industrial applications. However, in order to improve the control performance of the scalar control method, an encoder or speed tachometer is required, which makes it an expensive and less reliable solution.

Nowadays various modifications of vector control are applied to get a high quality control of induction motor drives. They enable to control the induction motor as a separately excited DC motor [6] – [9]. This type of control usually requires transformation of induction motor variables into a rotating system, which implies of using powerful computational means (digital signal processors) for its implementation.

Various sensorless methods for rotor position of electrical drives with induction motors [10] – [13] have also been proposed for decades with the aim of overcoming the disadvantages of hardware complexity and cost, which at the same time are able to ensure higher mechanical robustness and reliability.

Also sliding mode control presents an effective and robust method of controlling nonlinear systems [14] – [17] which can be employed in position and speed control of an AC drive. One of the recent very popular and promising methods for solving control tasks in various nonlinear systems are soft computing methods which include fuzzy logic, artificial neural networks, evolutionary algorithms and their combinations. They have already been applied in induction motor drive control [18] – [23]. However, compared to the classic control theory the intelligent controllers usually are very complex; they require powerful calculation processes, especially in real-time implementation and, moreover, the fuzzy control theory does not pay much attention to the stability of the system.

From the above it is clear that the effort of ensuring a high quality of induction motor control involves to take into account a nonlinear oscillating character of its dynamics, the changes of parameters, the influence of external disturbances and simple implementation of the control algorithm. All this factors makes development of advanced induction machine control still a challenging task.

The paper deals with the design and verification of a stable control structure with a parallel model which is applied in speed control of a drive with an AC motor. In comparison with vector control structures, the proposed control structure is considerably simpler, linear, strongly stable, robust, and it shows identical dynamical properties.

I. DESIGN OF MODEL REFERENCE CONTROL STRUCTURE

The desired behaviour of a controlled system (i.e. the system dynamics with which the desired values of its output and possibly also of state variables are to be obtained) is very often described by a reference model. It is suitable to choose this model as a linear system, because then it can be optimally designed using standard methods of optimal control theory.

Let us select the reference model for the controlled system in the form of the state-space description

$$\frac{\mathrm{d}\mathbf{x}_\mathrm{M}}{\mathrm{d}t} = \mathbf{A}_\mathrm{M}\mathbf{x}_\mathrm{M} + \mathbf{B}_\mathrm{M}\mathbf{w}, \tag{1}$$

where $\mathbf{x}_\mathrm{M}$ is the state vector of the reference model, $\mathbf{A}_\mathrm{M}$ is the state matrix of the reference model, $\mathbf{B}_\mathrm{M}$ is the input matrix of the model and $\mathbf{w}$ is the vector of the desired values.

The controlled structure is described in state space as a nonlinear continuous system with parametric as well as with additive disturbances (or „deviations“ from the reference model) in the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = (\mathbf{A}_\mathrm{M} + \Delta\mathbf{A})\mathbf{x} + (\mathbf{B}_\mathrm{M} + \Delta\mathbf{B})\mathbf{u} + \mathbf{v} =$$

$$= \mathbf{A}_\mathrm{M}\mathbf{x} + \mathbf{B}_\mathrm{M}\mathbf{u} + (\Delta\mathbf{A}\mathbf{x} + \Delta\mathbf{B}\mathbf{u} + \mathbf{v}), \tag{2}$$

where $\mathbf{x}$ is the state vector of the controlled structure, $\mathbf{u}$ is its input variable, $\Delta\mathbf{A}$, $\Delta\mathbf{B}$ are the matrices of the parametric disturbances and $\mathbf{v}$ is the vector of additive disturbances. Let us connect the reference model with the controlled system in accordance with the following relation, where $\mathbf{e}$ is the vector of the deviation of state variables of the model (1) and of the system (2)

$$\mathbf{e} = \mathbf{x}_\mathrm{M} - \mathbf{x}. \tag{3}$$

This expanded system can then be described regarding to the vector $\mathbf{e}$ by the equation

$$\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}t} = \mathbf{A}_\mathrm{M}\mathbf{e} - \mathbf{B}_\mathrm{M}\mathbf{u} + \mathbf{f}, \tag{4}$$

$$\mathbf{f}(\mathbf{x}, t) = -\Delta\mathbf{A}\mathbf{x} - \Delta\mathbf{B}\mathbf{u} - \mathbf{v} + \mathbf{B}_\mathrm{M}\mathbf{w}. \tag{5}$$

In (4), $\mathbf{f}$ is the generalised vector of the parametric as well as additive disturbances that we wish to be compensated by the vector $\mathbf{u}$.

In order to investigate the system´s asymptotic stability of (4) in terms of Lyapunov, let us use the positive definite Lyapunov function in the form

$$\mathbf{V} = \mathbf{e}^\mathrm{T}\mathbf{P}\mathbf{e}, \tag{6}$$

the derivation of which in terms of (1), (2), (3) and (4) after simple modifications is as follows

$$\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = \mathbf{e}^\mathrm{T}\left(\mathbf{A}_\mathrm{M}^\mathrm{T}\mathbf{P} + \mathbf{P}\mathbf{A}_\mathrm{M}\right)\mathbf{e} + 2(\mathbf{f}^\mathrm{T}\mathbf{z} - \mathbf{B}_\mathrm{M}\mathbf{u}^\mathrm{T}\mathbf{z}) =$$

$$= \mathbf{e}^\mathrm{T}\mathbf{Q}\mathbf{e} + 2(\mathbf{f}^\mathrm{T}\mathbf{z} - \mathbf{B}_\mathrm{M}\mathbf{u}^\mathrm{T}\mathbf{z}), \tag{7}$$

where the vector $\mathbf{z}$ is given as

$$\mathbf{z} = \mathbf{P}\mathbf{e}. \tag{8}$$

In (6), (7), (8) $\mathbf{P}$ is the positive definite matrix which satisfies the Lyapunov matrix equation

$$\mathbf{A}_\mathrm{M}^\mathrm{T}\mathbf{P} + \mathbf{P}\mathbf{A}_\mathrm{M} = -\mathbf{Q}. \tag{9}$$

According to the optimal control theory [24] for the state matrix of the reference model AM in the controllability form it is possible to use the positive definite matrix $\mathbf{Q} = -\alpha\,\mathbf{P}$.

The system (4) will be asymptotically stable if the derivation of Lyapunov function (6) is negative definite. If the control law of the controlled system is chosen in the form

$$\mathbf{u} = \mathbf{K}^\mathrm{T}\mathbf{z}, \tag{10}$$

and then (7) will be negative definite at the condition that

$$|z_i| \geq \frac{|f_i|}{K_i} \quad i = 1, 2, \ldots, n. \tag{11}$$

In order to achieve zero control deviation of the output variable in steady state, we will choose the first component of vector $\mathbf{e}$ to be the integral from the difference between the output variable of the reference model and the controlled system

$$e_1 = \int (y_M - y)\, dt. \tag{12}$$

If we assume the maximum values of the individual components of the disturbance vector $|f_i|$ are physically limited, then it is possible to ensure the condition (11) via relevant increase of values of the optional vector $\mathbf{K}$.

The block diagram of the designed controlled system is shown in Fig. 1.

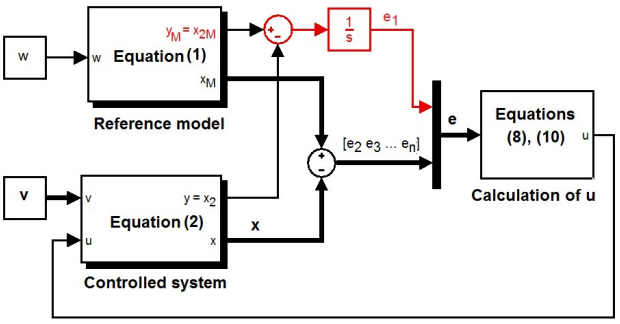


Fig. 1. Block diagram of the proposed controlled system.

The following basic features result from design of the above mentioned control structure:

- The desired dynamical performance of the controlled system is prescribed by the reference model (1), which in fact presents a stable linear system and thus it can always be designed to be optimal in terms of the chosen criteria, using control theory standard methods the for linear systems. Because the reference model prescribes the desired performance dynamically, during the whole transient state of the control process, it allows setting of an optimal motion dynamics of the controlled system.
- If the reference model is chosen optimally according to standard criteria of the optimization theory (based on the Riccati equation), then we are able to set its dynamics using a single parameter (parameter α) and by this the matrix $\mathbf{P}$ is defined at the same time such, that satisfies the Lyapunov matrix equation (9), which does not need to be solved, see [24].
- Control stability is ensured by condition (11), in which the optional parameter is the vector of gains $\mathbf{K}^\mathrm{T}$, or its individual positive components. It can be assumed that in technical systems the size of the individual components of the disturbance vector $\mathbf{f}$ is physically limited, and so, at

a sufficient size of components of the vector **K**, it is always possible to achieve zero control deviation of the output (controlled) variable in steady state.

II. Induction Motor and its Analytical Model

A squirrel-cage induction motor is described by stator currents and rotor fluxes by a current-flux model [9] in the {x, y} reference system, where the equations are as follows

$$\frac{d\psi_{2x}}{dt} = -\omega_g \psi_{2x} + M\omega_g i_{1x} + (\omega_1 - \omega)\psi_{2y}, \quad (13)$$

$$\frac{d\psi_{2y}}{dt} = -\omega_g \psi_{2y} + M\omega_g i_{1y} - (\omega_k - \omega)\psi_{2x}, \quad (14)$$

$$m = M\frac{p}{L1}\left(\psi_{2x} i_{1y} - \psi_{2y} i_{1x}\right), \quad (15)$$

$$\frac{J}{p}\frac{d\omega}{dt} = m - m_Z, \quad (16)$$

where, for the sake of simplification, the following constants have been introduced

$$\omega_g = R_2/L_2, \quad (17)$$

$$M = L_h. \quad (18)$$

and the used mathematical symbols are:

i_{1x}, i_{1y}	components of stator current space vector i_x
ω	motor mechanical angular speed
ω_1	angular speed of the stator voltage
ω_2	slip angular speed $\omega_2 = \omega_1 - \omega_m$
R_2	rotor phase resistance
Ψ_{2x}, Ψ_{2y}	stator and rotor magnetic flux
L_h, M	main inductance
L_1, L_2	leakage inductance
m	motor torque
p	number of pole pairs
J	moment of inertia
m_z	load torque

It is obvious from (13) – (18) that the induction motor presents a strongly nonlinear higher-order controlled system with oscillating character (in case of induction motor current-flux model it presents a 3rd order system).

For illustration, Fig. 2 shows the motor torque and angular speed responses at step change of motor inputs, with power supply from a current converter having a built-in current loop that compensates the circuit time delays.

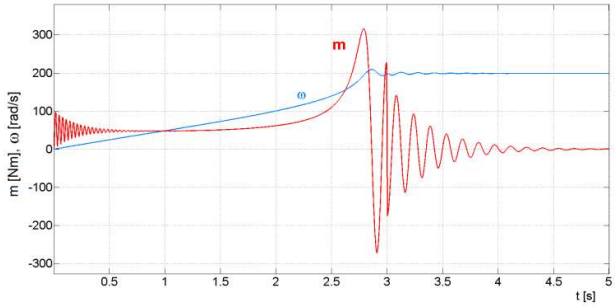


Fig. 2. Time responses of mechanics characteristics of the motor model at suplying by current vector.

Let's note that according to (13) – (18) the motor model is supplied by current where, in this case, the stator current vector was set at i_{1y} = 15 A and the angular speed reaches the value of ω_1 = 200 rad/s. The induction motor parameter values are specified in the Appendix.

In the systems of this type it is not suitable to design the control law using methods valid for linear systems. For this reason for control of the angular speed of the AC drive with induction motor we apply the control structure derived in the Chapter II.

III. Design of Control for Induction Motor Drive

The objective of AC induction motor drive control is controlling its angular speed, the dynamical behaviour of which is prescribed by the reference model.

As in our case the induction motor drive presents the 3rd order system, according to the optimization theory [24] for the reference model we can choose a system with matrices in the form

$$\mathbf{A}_M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{\alpha^3}{2} & -\frac{3\alpha^2}{2} & -\frac{3\alpha}{2} \end{bmatrix} \quad \mathbf{B}_M = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \quad (19)$$

According to [24] the optimal matrices **P** and **Q** for this system satisfying the Lyapunov matrix equation (9) are as follows

$$\mathbf{P} = \begin{bmatrix} \frac{\alpha^5}{2} & \alpha^4 & \frac{\alpha^3}{2} \\ \alpha^4 & \frac{5\alpha^3}{2} & \frac{3\alpha^2}{2} \\ \frac{\alpha^3}{2} & \frac{3\alpha^2}{2} & \frac{3\alpha}{2} \end{bmatrix} \quad \mathbf{Q} = \alpha \begin{bmatrix} \frac{\alpha^5}{2} & \alpha^4 & \frac{\alpha^3}{2} \\ \alpha^4 & \frac{5\alpha^3}{2} & \frac{3\alpha^2}{2} \\ \frac{\alpha^3}{2} & \frac{3\alpha^2}{2} & \frac{3\alpha}{2} \end{bmatrix} = \alpha\mathbf{P}. \quad (20)$$

In this model, the desired dynamics can be conveniently set by a single parameter α, and quality of the dynamics does not change; there are only changes in the settling time of the state variables or of the output variable of the reference model. The settling time for this model is equal approximately to $T_s \approx 5/\alpha$. Let's select the state variable of the controlled system (of the induction motor) $y = x_2 = \omega$ (the rotor angular speed) and $x_3 = dx_2/dt = \varepsilon$ (acceleration of the motor on the shaft that corresponds to its dynamic moment m). From the time responses in Fig. 2 it follows that the concrete model of the induction motor starts up in about 2.5 sec, so let us choose for the model dynamics the value α = 5. Let the optional vector $\mathbf{K}^T$ is equal to [0.001 0.001 0.001].

The resulting control scheme for controlling the angular speed of the induction motor drive in accordance with the derived control structure presented in Fig. 1 is shown in Fig. 3.

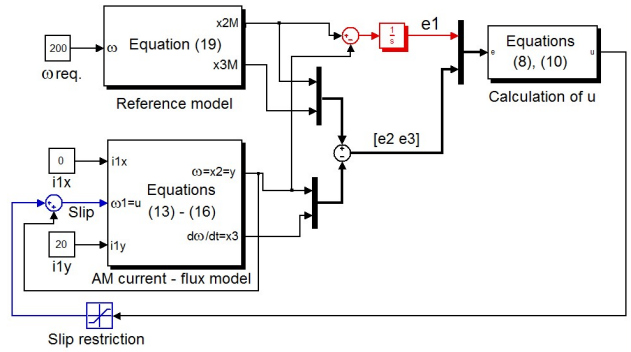


Fig. 3. Block diagram of induction motor control.

In order the induction motor would operate in the stable part of its torque characteristics, the diagram in Fig. 3 has been completed by conversion of the actuating signal into the slip, and also its limitation has been added. This, however, does not imply any changes in general character of the structure presented in Fig. 1.

The limitation of magnitudes of the state variables of the controlled system (current, torque) can be realized in the reference model without affecting stability of the control loop (because the reference model is linear and the poles of the linear system are not changed with the limitations).

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-16-0206.

I also wish to thank Mrs. Professor Daniela Perduková and Mr. Professor Pavol Fedor for their guidance so far.

REFERENCES

[1] D.W. Novotny, T.A. Lipo, *Vector control and dynamics of AC drives*, Oxford: Clarendon Press, UK, 1996.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] S. J. Dodds, *Feedback control: linear, nonlinear and robust techniques and design with industrial applications*, London: Springer, UK, 2015.

[3] J. Boldea, A.S. Nasar, *Vector control of AC drives*, London: CRC Press, UK, 1992.

[4] J. Boldea, A.S. Nasar, *Scalar V/f and I-f Control of AC Motor Drives: An overview*, Electric drives, 3rd ed., 2017, pp. 571-593J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

[5] N. Barsoum, D. Faiman, P. Vasant, „Scalar control on speed drive for AC motor," in: *Proc. of the sixth Global conference on power control and optimization*, Vol. 1499, 2012, pp. 169-174.

[6] P. Vas, Vector control of AC machines, New York: Clarendon, 1990

[7] A.K.Chattopadhyay, „Advances in vector control of AC motor drives - A review," in: *Sadhana-academy proceedings in Engineering sciences*, Vol. 22, 1997, pp. 797-820.

[8] Z. Zhang, Z.G. Liu, L.T. Zhao, L.J. Diao, et.al., „Research on AC drive test system based on vector control," in: *Proceedings of the 2013 International conference on Electrical and Information Technologies for Rail Transportation* (EITRT2013), Vol. 288, 2014, pp. 537-546

[9] P. Brandstetter, M. Kuchar, H.H. Vo, C.S.T. Dong, „Induction motor drive with PWM direct torque control," in: *Proceedings of the 2017 18th International Scientific Conference on Electric Power Engineering-EPE*, 2017, art. no. 7967268. DOI: 10.1109/EPE.2017.7967268.

[10] T. Michalski, C. Lopez, A. Garcia, L. Romeral, „Sensorless Control of Five Phase PMSM Based on Extended Kalman Filter," in: *Proc. of the IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 2016, pp. 2904-2909.

[11] S. Shinnaka, „Sensorless vector control of induction motors using minimum dimensional flux D-state observer with instantaneous speed estimation," *Electrical Engineering in Japan*, Volume 196, Issue 1, 2016, pp. 31-41.

[12] P. Makys, G. Asher, M. Sumner, Q. Gao, J. Vittek, „Shaft sensorless speed control of induction motor drive," in: Proc. of 12th International Power Electronics and Motion Control Conference, Portoroz, Slovenia, Vol. 1-4, 2006, pp. 1686-+ .

[13] P. Brandstetter, M. Dobrovsky, O. Petrtyl, C.S.T. Dong, H.H. Vo, „Sensorless control of induction motor drive using BEMF-MRAS with Kalman filter," in: *Proc. of 11th International Conference on ELEKTRO 2016*, High Tatras; Slovakia, 2016, pp. 121-126.

[14] S.J. Dodds, J. Vittek, „Sliding Mode Control of PMSM Drives Subject to Torsion Oscillations in the Mechanical Load, „ in: *Proc. of 13th International Power Electronics and Motion Control Conference EPE PEMC*, Poznan, Poland, 2008, Vols. 1-5, pp. 2551-+.

[15] R.H. Kumar, A. Iqbal, N.CH. Lenin, „Review of recent advancements of direct torque control in induction motor drives - a decade of progress," *IET Power Electronics*, Vol. 11, No. 1, 2018, pp. 1-15.

[16] R. Kamaleshwar, A. Prahadeeshwaran, N. Srivats Athindran, „Sensorless sliding mode control of SVPWM fed induction motor drive using DSS speed estimator," in: *Proc. of International Conference on Innovative Research in Electrical Sciences* (IICIRES2017), India, 2017, WOS:000425228200019.

[17] A. Oukaci, R. Toufouti, D. Dib, L. Atarsia, „Comparison performance between sliding mode control and nonlinear control, application to induction motor, *Electrical Engineering*, Vol. 99, No. 1, 2017, pp. 33-45.

[18] P. Brandstetter, M. Kuchar, I. Neborak, "Selected Applications of Artificial Neural Networks in the Control of AC Induction Motor Drives," *International Review on Modelling and Simulations*, Vol. 4, No. 3, 2011, pp. 1084-1093.

[19] P. Girovsky, J. Timko, J. Zilkova, "Shaft Sensor-less FOC Control of an Induction Motor Using Neural Estimators," *Acta Polytechnica Hungarica*, Vol. 9, No. 4, 2012, pp. 31-45.

[20] P. Palacky, P. Hudecek, A. Havel, "Real-Time Estimation of Induction Motor Parameters Based on the Genetic Algorithm," in: *International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, Book Series: Advances in Intelligent Systems and Computing*, vol. 189, 2013, pp. 401-409.

[21] T. Radicova, M. Zalman, "Master Slave LMPM Position Control Using Genetic Algorithms," *Soft Computing Models in Industrial and Environmental Applications, Book Series: Advances in Intelligent Systems and Computing*, vol. 188, 2013, pp. 73-82.

[22] L. Qin, Y. Jiang, X.X. Tan, L.Y. Li, J.Y. Zhao, „Application Research of Fuzzy Control for AC Motor on the PLC Platform," *Advances in Civil Engineering*, PTS 1-6, Vol. 255-260, 2011, pp. 2136-+.

[23] M. Strefezza, H. Kobayashi, K. Fujikawa, Y. Dote, „Fuzzy Robust-Control for AC Drive System," *Journal of Systems Engineering*, Vol. 4, No. 2, 1994, pp. 87-96.

[24] V. D. Furasov, *Ustojčivosť dviženija, ocenki i stabilizacija*. Moskva: Nauka, 1977, pp. 213-220.

[25] P. Talian, D. Perduková, P. Fedor, „Stable and Robust Tension Controller for Middle Section of Continuous Line," *Elektronika Ir Elektrotechnika*, Vol. 24, No. 1, 2018, pp. 3-10.

Superconducting properties of strongly disordered molybdenum nitride ultrathin films in the vicinity of superconductor-insulator transition

[1]*Marek KUZMIAK (2nd year),*
Supervisor: [2]*Pavol SZABÓ*

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic
[1,2]Centre of Low Temperature Physics, Institute of Experimental Physics of Košice, Slovak Academy of Sciences, Slovak Republic

[1]marek.kuzmiak@tuke.sk, [2]pszabo@saske.sk

***Abstract*—The subject of my PhD thesis is the study of the influence of low dimensionality and nano-size effects on superconducting properties of ultrathin films. I study physical properties of highly disordered superconducting ultrathin MoN films near the superconductor-insulator transition (SIT). The disorder is introduced into the samples by reducing their thickness. In this article, I present our experimental results, obtained via transport and low temperature STM/STS measurement in highly disordered 3 nm ultrathin films, which are near the critical disorder in the vicinity of the SIT. The inhomogeneous superconducting properties observed on the surfaces of the studied thin films are rather associated with locally different disorder than with the Bosonic localization of the superconducting condensate. Our results indicate the presence of Fermionic localization and the closeness of Fermionic SIT in this system.**

***Keywords*—superconductivity, disordered system, scanning tunneling microscopy (STM), scanning tunneling spectroscopy (STS), superconductor-insulator transition (SIT), molybdenum nitride (MoN)**

I. Introduction

The superconducting state is characterized by the wave function $\psi(\vec{r}) = \Delta e^{i\Theta(\vec{r})}$, with the amplitude Δ and phase Θ. The increase of disorder can destroy superconductivity in two different ways, either by suppressing the amplitude Δ (Fermionic scenario) [1] or by breaking the phase coherence Θ (Bosonic scenario) [2]. In the case of Fermionic scenario the increased disorder due to enhanced Coulomb interactions breaks Cooper pairs into fermionic states, which leads to the suppression of superconducting properties. The bosonic mechanism acts through fluctuating phase coherence. While the Fermionic scenario predicts homogeneous superconducting properties, in the Bosonic scenario due to the phase fluctuations the Cooper pairs start to form locally fluctuating superconducting islands (in the range of the superconducting coherence length), which can persist as phase incoherent superconducting areas even in the (insulating) normal state.

In this paper we study the superconducting properties in strongly disordered MoN ultrathin films, where the disorder is introduced through reducing thickness. Our preliminary results show [3], that our 3 nm thin samples studied in this paper are near the critical disorder, where the SIT occurs. Applying low temparature STM/STS measurements we show, that the observed local superconducting inhomogeneities are related to the variation of disorder in the highly corrugated surfaces and not the bosonic effects.

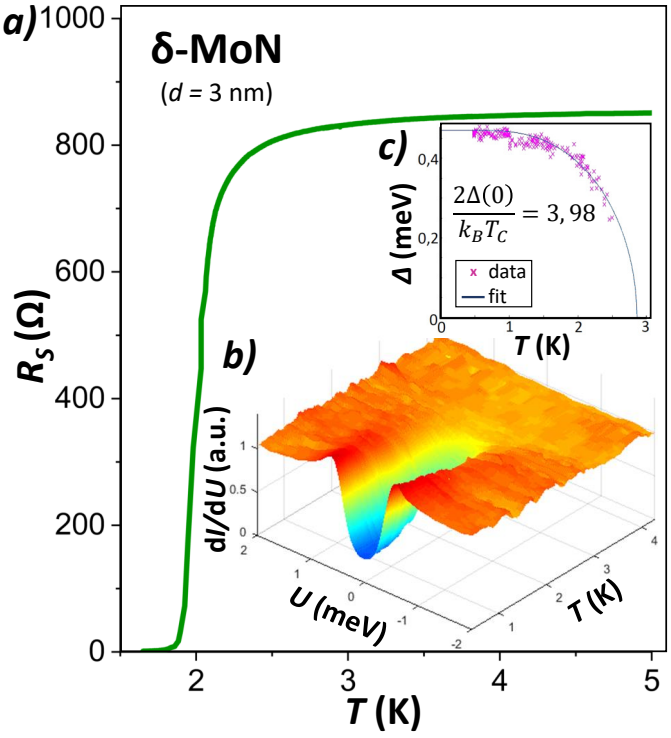


Fig. 1. The temperature dependence of sheet resistance R_S of 3 nm ultrathin δ-MoN film near T_C is shown in the main panel with green line (a). The lower inset (b) is a 3D plot of the typical temperature dependence of locally measured STM differential conductance spectra. All spectra were normalized to their values in the normal state at 2 meV bias energy. The upper inset (c) shows the temperature dependence of the superconducting energy gap (X-symbols) obtained from fitting the tunnelling spectra from (b) to the BCS model. The solid blue line shows the predictions of the BCS theory for the strongly coupled superconductor with the coupling strength $2\Delta(0)/k_B T_C = 3.98$.

II. Experiment

The MoN ultrathin film of 3 nm thickness was prepared by reactive magnetron sputtering onto a sapphire substrate

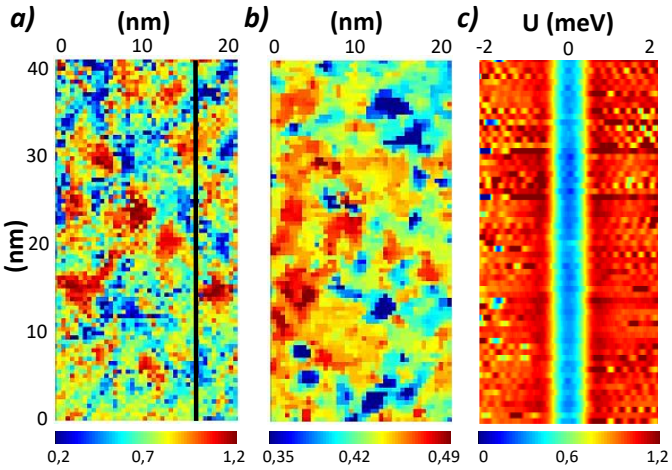


Fig. 2. (a) STM surface topography of a 20×40 nm^2 surface area at 500 mK. (b) superconducting energy gap map obtained from tunneling spectra measured at each point of the topography from (a). The top view of the tunneling spectra measured along a 40 nm long line plotted in Fig. (a) in a voltage window ± 2 meV in the marked colour scale is shown in (c).

in an argon-nitrogen mixture [4]. The crystallographic structure of MoN ultrathin film was characterized by XRD measurements. We found that our film has a hexagonal structure. It is a stoichiometric phase of δ-MoN. Transport measurements of the sheet resistance R_S were performed using the Van der Pauw four-probe technique [5]. The STM/STS measurements were carried out via a sub-Kelvin STM system developed in the Centre of Low Temperature Physics in Košice. The value of the upper critical magnetic field determined from suppression of the superconducting energy gap is $H_{C2} = 6$ T, the calculated coherence length is $\xi = 7.41$ nm [3].

III. Results and Discussion

Fig. 1 a) shows the temperature dependence of the sheet resistance $R_S(T)$ of a 3 nm ultrathin δ-MoN film near the transition temperature T_C. The sheet resistance starts to drop at $T = 3.5$ K and the zero resistance is obtained at about $T = 1.8$ K. The temperature dependence of the sheet resistance shows a single-phase superconducting transition width $\Delta T_C/T_C \approx 0.16$, where ΔT_C is the difference in T_C determined at 10 % and 90 % of the normal state sheet resistance $R_N \approx 830\ \Omega$.

Fig. 1 b) represents the 3-dimensional (3D) plot of a typical temperature dependence of the STM differential conductance spectra in zero magnetic field. All measured spectra were normalized to their normal state values at $U = 2$ meV. As we can see, the tunneling spectra reveal evident superconducting gap structure with symmetric maxima positioned at the gap energies. The gap structure is strongly suppressed with increased temperature and totally disappears at the critical temperature at around 3 K. The precise value of the superconducting gap and the critical temperature can be determined from fitting of the temperature dependence of the tunnelling spectra. It is well known, that the tunneling conductance, measured between a superconductor and a normal metal is directly proportional to the superconducting density of states (DOS), defined in the BCS theory as $N_S(E) = E/\sqrt{E^2 - \Delta^2}$, where E is energy and Δ is the superconducting energy gap. This DOS function taking into account the finite temperature and the additional spectral smearing effects defined as $E = E + i\Gamma$ (Γ is the spectral smearing parameter) enables precise determination of the energy gap Δ values at different temperatures [6]. Fig. 1 c) shows the obtained $\Delta(T)$ dependence (X-symbols) and the theoretical predictions of the BCS theory for superconductors with the coupling strength $2\Delta(0)/k_B T_C = 3.98$ [7] (blue line). It is evident, that the $\Delta(T)$ values obtained from fitting follow the BCS curve for strongly coupled superconductors.

The surface of the studied MoN ultrathin film was investigated by means of STM topography measurements. Fig. 2 a) shows images of a 20×40 nm^2 surface area measured at 500 mK. This surface reveals well defined yellow-red $30 - 100$ nm^2 areas protruding 0.8 - 1.2 nm above the deepest green-blue regions. We performed local STS measurements in each point of the surface topography shown in Fig. 2 a). The local value of the superconducting energy gap was determined from fitting the locally measured tunneling spectra to the BCS DOS, as described in the text above. The superconducting gap map of the surface in Fig. 2 a) is constructed in Fig. 2 b). Fig. 2 c) plots the top view of the evolution of the STM spectra measured at 500 mK along the black line in Fig. 2 a).

The almost continuous contrast lines between different colours of Figure Fig. 2 c) which characterize the superconducting energy gap indicate homogeneous superconducting properties. However, the gap map, shown in Fig. 2 b) reveals evident gap inhomogeneities in well-defined areas. Comparing figures Fig. 2 a) and Fig. 2 b), it's evident, that the local values of the energy gap are highly sensitive to the surface corrugations. Δ has approx. 20 % bigger values in protruding red surface areas than in the deeper blue regions. Hereby, these areas with different gap values cannot be formed through Bosonic localization. Similar variation of Δ has been observed in ultrathin MoC thin films [8], where the correlation of the Δ inhomogeneities with the surface corrugation has been explained with differences in the local disorder in thinner and thicker parts of the thin film. It has been shown, that the thicker parts of the samples reveal lower disorder, which causes a weaker suppression of superconductivity and higher values of Δ. Comparing the conclusions of the paper [8] with our data, it's evident that the local variation of Δ in our MoN films is due to the change of their thickness and the superconducting state remains spatially homogeneous in all flat sample areas with constant thickness. This excludes the Bosonic origin of the local variation of the superconducting energy gap and strongly supports the Fermionic localization of quasiparticles in our 3 nm ultrathin MoN films.

IV. Conclusion

This paper is dealing with the experimental study of the superconducting properties in strongly disordered 3 nm ultrathin MoN films, which are near the critical disorder in the vicinity of the superconductor-insulator transition. Our low temperature STM/STS data detected local inhomogeneities of the superconducting energy gap Δ, which can be correlated with the corrugation of the studied surfaces. Due to lower disorder, the superconducting properties in thicker areas of the thin films are less suppressed and the energy gap reveals higher values. This suggests, that the observed islands with locally different values of the superconducting energy gap are not related to Bosonic localization. The increase of disorder in MoN sam-

ples suppresses superconductivity through Fermionic mechanism. These results indicate the closeness of the Fermionic suprconductor-insulator transition in this system.

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 contracts No. VEGA 1/0743/19, VEGA 2/0058/20, APVV-18-0358, and by the U.S. Steel Košice.

REFERENCES

[1] A. M. Finkel'stein, "Suppression of superconductivity in homogeneously disordered systems," *Physica B: Condensed Matter*, vol. 197, pp. 636–648, March 1994.

[2] M. P. A. Fisher and et al., "Presence of quantum diffusion in two dimensions: Universal resistance at the superconductor-insulator transition," *Physical Review Letters*, vol. 64, pp. 587–590, January 1990.

[3] M. Kuzmiak and et al., *to be published*.

[4] M. Kožejova and et al., "Evaluation of hydrogen evolution reaction activity of molybdenum nitride thin films on their nitrogen content," *Electrochimica Acta*, vol. 315, pp. 9–16, August 2019.

[5] L. J. V. der Pauw, "A method of measuring specific resistivity and hall effect of discs of arbitrary shape," *Philips Research Reports*, vol. 13, pp. 1–9, February 1958.

[6] R. C. Dynes and et al., "Direct Measurement of Quasiparticle-Lifetime Broadening in a Strong-Coupled Superconductor," *Physical Review Letters*, vol. 41, pp. 1509–1512, August 1978.

[7] M. Tinkham, *Introduction to superconductivity (2nd ed.)*. McGraw Hill Book Co., NY, USA: University Science Books, 2004.

[8] P. Szabó and et al., "Fermionic scenario for the destruction of superconductivity in ultrathin MoC films evidenced by STM measurements," *Physical Review B*, vol. 93, p. 014505, January 2016.

Robust 3D Facial Expression Recognition

[1]*Fouzia Adjailia (3rd year),*
Supervisor: [2]*Peter Sinčák*

[1,2]Dep. Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]fouzia.adjailia@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—In this paper, we propose a facial expression recognition approach where we combined features extracted from 2D textured images and 3D geometric data using Local Binary Pattern (LBP) and 3D Voxel Histogram of Oriented Gradients (3DVHOG) respectively. We carried out some pre-processing methods on the MDPA-FACE3D and Bosphorus datasets. The features were used to classify faces into seven universal emotions, namely anger, disgust, fear, happiness, sadness, neutral, and surprise. Using Support Vector Machine classifier, we achieved the accuracy of 88.5% and 92.9% on the MDPA-FACE3D and the Bosphorus datasets, respectively.**

***Keywords*—facial expression recognition, histogram of oriented gradient, local binary pattern, descriptors, feature extraction**

I. Introduction

As human beings, we use many channels to communicate our thoughts and emotions; verbal, gesture of gait, language of the body or facial expressions. 3D images can be computationally intensive to work with in comparison to 2D images. However, they have a more accurate and comprehensive description of the face. They give the parameter of face depth, which is important for the recognition of face and facial expression because muscles can not be easily seen with 2D images such as a concave or convex facial structure. For facial expression analysis, the muscle activity of the face may significantly help to understand what emotion is conveyed, and these muscular movements can be captured in detail by 3D imaging. In order to change all the 2D image processing problems of technology that is shifting facial imaging to other dimensions, the 3D mapping will be addressed. For 3D imaging, variations in lighting are not a problem, since the mask is not deformed and the geometric data does not differ based on the lighting.

II. Related Studies

Until around the early 2000s, algorithms integrating effects from 2D and 3D data did not exist. Nowadays, the most direct techniques in this field use the combination of characteristics acquired independently of bi-dimensional or three-dimensional methods, such as texture details, position of landmarks, facial forms, and curvature, to identify expressions and calculate their intensity. Experiments show that integrating features obtained in various modalities helps to capture the general features of facial deformation and increases the precision of identification. The topic of facial expression recognition was explored by berretti2010set. To this end, an initial method is suggested that measures scale-invariant feature transform (SIFT) descriptors on a range of depth image facial landmarks, and then selects the sub-set of most the important characteristics. An overall identification rate of 77.5% on the BU-3DFE database was obtained using SVM classification of the chosen features. Comparative assessment of a typical experimental setup illustrates that state-of-the-art outcomes can be obtained from their approach.

III. Datasets Description

A. Bosphorus

The Bosphorus database [1] was also used for scientific purposes by Savran et al. A total of 4,666 scans obtained from 105 subjects are included in the collection, 61 of which are male and 44 are female. Included are some facial expressions that are expressed in two ways, first being the fundamental expressions of happiness, surprise, fear, sadness, anger, disgust . The other are Action Units-based expressions. For each word, each subject will preferably include a single frontal face picture and a 3D landmark file,

B. IMPA-FACE3D

To assess the analysis of facial expression recognition, in 2008, the database IMPA-FACE3D was developed, in particular. The neutral face (face with the location of the front camera -0 degrees and without facial expression) and the six universal expressions suggested by Ekman amongst human races are the basis for this purpose: happy, sad, surprise, anger, disgust, and fear. Overall, the data collection consists of 22 men and 16 women, the bulk of which are between 20 and 50 years of age. For all entities, 14 samples were acquired, summarizing 532 samples.

IV. Proposed Approach

A. Experimental Setup

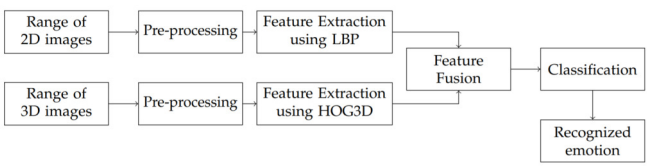


Fig. 1. Pipeline of the proposed approach

Figure 1 represents the overall architecture of the proposed approach, it consists of three parts, image pre-processing of both 2D and 3D input images, as well as the feature extraction phase, where Histogram of oriented gradients (HOG3D) and Local Binary Patterns (LBP) used on 3D and 2D images respectively to extract features, the two features vectors are

concatenated, the integration of the two feature vectors are fed to different classifiers to assess the facial expression classification process.

B. Data pre-processing

Table I presents the transformation approaches applied to the Bosphorus and IMPA-FACE3D datasets. Figure 2 represents an example of the pre-processing phases applied for the 3D images from the IMPA-FACE dataset.

TABLE I
TRANSFORMATION METHODS APPLIED TO BOSPHORUS AND IMPA-FACE3D

dataset	Bosphorus		IMPA-FACE3D	
	2D	3D	2D	3D
write Point cloud files		x		
face cropping			x	
grey scale	x		x	
uniform down-sampling		x		x
region of interest		x		x
occupancy grid		x		x

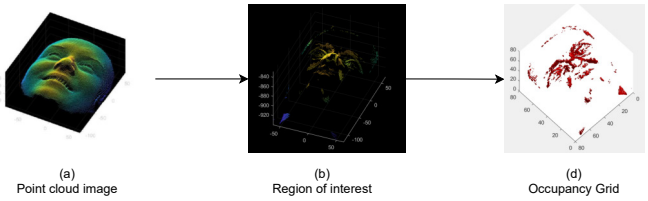


Fig. 2. Pre-processing methods applied in our experiment on IMPA-FACE3D dataset

C. Feature Extraction

3D Voxel HOG [2] is based on Dalal and Triggs' original Histogram of Oriented Gradients [3] By using voxels over pixels, it extends the approach and stretches the original histograms into 2 dimensions. This implementation was developed for the identification of local object structures, for use in a context for risk analysis in which it is used to define the object's risk-related properties figure 3

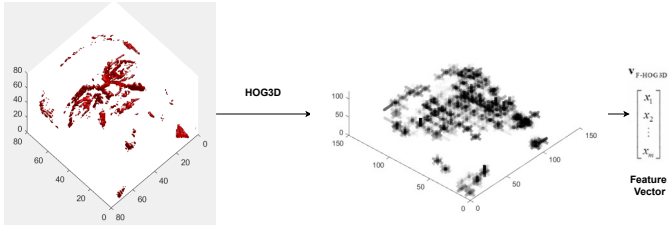


Fig. 3. Feature Extraction from the 3D images

Figure 4 presents the local binary pattern extraction from Bosphorus dataset where the number of neighbors is 8.

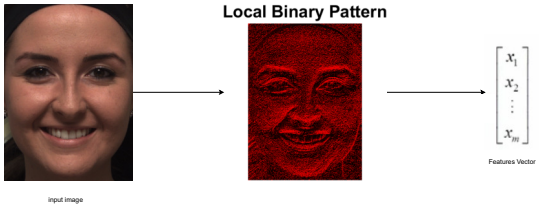


Fig. 4. Feature extraction from the 2D images using LBP

V. Discussion

Tables II reports the confusion matrix and show the result obtained on the IMPA-FACE 3D dataset. We achieved a model accuracy of 88.5%.

TABLE II
CONFUSION MATRIX; PC: PREDICTED CLASSES; AC: ACTUAL CLASSES

PC/AC	anger	disgust	fear	joy	neutral	sadness	surprise
anger	29	7	0	0	0	0	0
disgust	0	28	8	0	0	0	0
fear	0	0	35	1	0	0	0
joy	0	0	1	34	1	0	0
neutral	0	0	0	1	35	0	0
sadness	0	0	0	0	0	36	0
surprise	0	0	0	0	0	10	26

Tables III presents the confusion matrix and show the result obtained on the Bosphorus dataset. We achieved a model accuracy of 92.9%.

TABLE III
CONFUSION MATRIX; PC: PREDICTED CLASSES; AC: ACTUAL CLASSES

PC/AC	anger	disgust	fear	joy	sadness	surprise
anger	103	3	0	0	0	0
disgust	0	62	4	0	0	0
fear	0	1	68	2	0	0
joy	0	0	3	68	1	0
sadness	0	0	0	7	62	0
surprise	0	0	0	0	12	58

The findings demonstrates that the proposed approach outperforms the results obtained in recent research, table IV

TABLE IV
COMPARISON IN TERMS OF CLASSIFICATION RATE

Reference	Database	Methodology	Expressions	Accuracy %
vretos20113d	Bosphorus	Zernike moments + SVM	6	60.00
chun2013facial	Bosphorus	LBP + SVM	6	76.98
wang2013learning	Bosphorus	LBP + SVM	6	76.56
azazi2015towards	Bosphorus	SURF + SVM with PE	7	79.00
azazi2015towards	Bosphorus	SURF + SVM with EPE	7	84.00
li2015efficient	BU–3DFE/Bosphorus	Multi-modal 2D and 3D descriptors	6/7	79.72
yang2015automatic	BU-3DFE	Geometric scattering representations	6	82.73
berretti2010set	BU-3DFE	scale-invariant feature transform (SIFT)	6	77.5
Our Approach	**Bosphorus**	**HOG3D+LBP+SVM**	**6**	**92.9**
Our Approach	**MPA-FACE 3D**	**HOG3D+LBP+SVM**	**7**	**88.5**

VI. Conclusion and Future Work

In this paper, we proposed a new approach to effectively recognize human facial expressions. We applied cutting edge techniques for 3D image pre-processing, We used Local Binary Patterns and 3D voxel Histogram of Oriented Gradients to extract features from 2D and 3D images respectively. Then we used a Support Vector Machine to assess the classification of the facial expression. We test the approach using the Bosphorus and the IMPA-FACE 3D dataset and obtained an accuracy of 92.9% and 88.5% respectively.

Acknowledgment

This research work was supported by APVV project 015-0730 "Cloud Based Human Robot Interaction" and FEI GRANT – 2021 – Increase server computing power to work with state-of-the-art artificial intelligence algorithms.

References

[1] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *European workshop on biometrics and identity management*. Springer, 2008, pp. 47–56.

[2] R. Dupre, V. Argyriou, D. Greenhill, and G. Tzimiropoulos, "A 3d scene analysis framework and descriptors for risk evaluation," in *2015 International Conference on 3D Vision*. IEEE, 2015, pp. 100–108.

[3] T. Q. Nguyen, S. H. Kim, and I. S. Na, "Fast pedestrian detection using histogram of oriented gradients and principal components analysis," *International Journal of Contents*, vol. 9, no. 3, pp. 1–8, 2013.

Environmental impacts on hybrid FSO/RF systems

[1]*Maroš LAPČÁK (1st year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]maros.lapcak@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*—This publication deals with the technical analysis of hybrid FSO / RF systems and the effects of the environment on individual lines. Since we want to achieve almost 100% reliability, it is important to know the parameters of the transmission environment for FSO and RF line. Each of them has its limitations and, as a result, a weather monitoring station is created. The FSO is most affected by fog and the concentration of particles in the air, while the RF line is affected by rain and snow. Therefore, it is important to record these conditions using a weather monitoring station to ensure smooth transmission for both lines.**

***Keywords*—FSO, RF, Hard Switching, weather conditions.**

I. Introduction

In recent decades, FSO (Free Space Optics) optical fiber less communication systems have become an important direction in the application of optoelectronic technologies. FSO is also known as fiber less photonics. To obtain a broadband communication channel, high frequency modulated light pulses are used to transmit data over the atmosphere [1], [2]. FSO is an optical communication technology in which data is transmitted by light propagation in free space, which allows optical connection. FSO works similarly to Optical Fiber Cable (OFC), but the only difference is that optical beams are emitted through air instead of OFC cores, which are glass fibers. The FSO system consists of an optical transmitter and a receiver at both ends, which provides two-way (full duplex) communication. FSO is a technology that for its proper functioning needs direct visibility of the so-called LOS (Line of Sight) [3], [4]. The use of lasers is a simple concept like optical transmission using cables from Optical fiber; the only difference is the transmission media. Light passes through air faster than glass, so FSO technology can be classified as optical communication at the speed of light. FSO communication is considered as an alternative to long distance radio communication (LOS) communication systems [5].

Figure 1 The principle of operation of the FSO system.

II. Theory

Visibility is an important and integral condition for the functioning of FSO systems. This is described as the transparency of the atmosphere estimated by the observer. The current value of visibility is useful due to the occurrence of different types of particles in the atmosphere (smog, fog, haze, clouds, etc.). Dense fog between the two FSO heads reduces visibility to a few meters, with haze or clouds having the same impact on FSO systems. Deteriorated or even low visibility reduces the overall efficiency and availability of FSO systems. It follows that attenuation and scattering are precisely in these conditions. The attenuation of the optical signal can reach several hundred dB/km in conditions of reduced visibility; specifically, for short wavelengths [6].

Commercial hybrid RF/FSO system performs well in temperate regions as reported in many literatures. Tropical region has different scenario where rain frequency and intensity are high. Since optics transmission uses light-wave that transmits particle-like electromagnetic signals, large particles floating in the open space block the transmission path. Commercial hybrid systems have self-mitigating mechanism by transmitting lower data for both RF and optics transmission channels to sustain high link reliability [7].

For FSO systems, there are several transmission windows that are almost transparent (attenuation <0.2 dB/km), in the wavelength range from 780 nm to 1600 nm. These windows are located around several specific wavelengths:

- 850 nm
- 1060 nm
- 1250 nm
- 1550 nm [5].

The installation of FSO heads within the premises of the Technical University in Košice ensures communication at a wavelength of 850 nm. The distance between the heads is approximately 250 m.

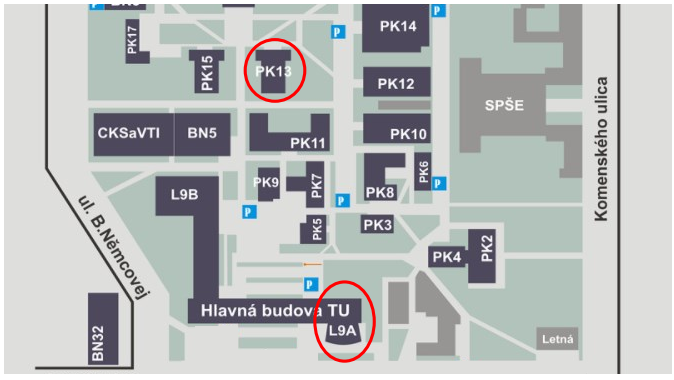


Figure 2 Location of FSO heads in the TUKE complex.

A. Atmospheric properties affecting the transmission of FSO systems

The FSO system enables the transmission of information by carrying an optical signal via an unregulated channel. FSO systems facilitate a secure way of communication without electromagnetic interference at high bandwidth. They are used in terrestrial and satellite connections. However, the main problem with the external FSO system is the random nature of the atmosphere, which causes a deterioration in connection performance. Turbulence caused by the formation of vortices of different sizes and refractive indices and weather conditions, such as fog, smog, rain, etc., can result in severe performance degradation [8]. The performance of the FSO system is affected by many factors such as:

- Absorption.
- Scatter
  - Rayleigh scatter.
  - Mie scatter.

The main parameters of the transmission environment that affect the transmission of the optical signal are mainly rain, fog, snow, turbulence and scintillation. These phenomena occurring in the atmosphere to some extent affect the optical beam, whether they attenuate or degrade it, which causes the FSO line to fail. Therefore, the availability and reliability of FSO systems is reduced. phenomena that affect the transmission environment are:

Rain

Rain is formed by water vapor contained in the atmosphere. It consists of water drops, the shape and number of which are variable in time and space. The scattering due to precipitation is called non-selective scattering because the radius of the raindrops (100-1000 μm) is significantly larger than the wavelength of typical FSO systems. The laser can pass through the raindrop particle, with less scattering. The turbidity particles are very small and remain longer in the atmosphere, but the rain particles are very large and remain shorter in the atmosphere. This is the main reason why rain attenuation is less than turbidity [5].

Fog

Fog significantly attenuates visible radiation. The optical beam of light is absorbed, scattered and reflected by an obstacle caused by fog. The scattering caused by fog, also known as Mie scattering, is largely a matter of increasing the transmitted power in FSO, the wavelength of the signal is chosen to operate in the low absorption bands. Therefore, the contribution of absorption to the overall attenuation coefficient is very small compared to the scattering effect. It is important to study the dispersion of fog particles in order to predict the attenuation in the planning and installation of a wireless network [5, 9].

Snow

Snow also affects the performance of FSO connections. The attenuation caused by snow varies depending on the location and atmospheric conditions at a particular location. The physical properties of snow are the main factor causing the degradation of the optical connection of the free space. The attenuation coefficient is given by:

$$\gamma = a.s^b \tag{1}$$

where γ represents the attenuation due to snow, s is the amount of snow and a, b are constants depending on the physical properties of snow [10]. Snow attenuation can be caused by wet or dry snow. The moisture or dryness of the snow determines the amount of liquid content in the falling snow. When the temperature in the air drops just above freezing, part of the snow melts and falls like wet snow. In dry snow, the temperature is completely below freezing, which means that the snow has a low liquid content [11].

Beam spreading

Beam propagation describes the expansion of the beam size in the target above the expected limit due to diffraction when the beam propagates in a turbulent atmosphere.

Turbulence

Turbulence can lead to degradation of the optical transmission beam. A change in the refractive index causes the refraction of the beam at different angles and the propagation of the optical beam occurs. Turbulence adversely affects the reliability and quality of the connection. There are short-term and long-term signal outages, attenuation. A partial solution to this problem is the hybridization of communication systems such as optical and radio frequency (FSO/RF). These hybrid FSO/RF are systems that operate on an optical line in good weather conditions and on a backup RF link during outages. Clean air turbulence phenomena affect the propagation of the optical beam by both spatial and temporal random fluctuations in the refractive index due to changes in temperature, pressure and wind along the optical propagation path. Atmospheric turbulence primarily causes phase shifts of propagating optical signals, leading to distortion in the front end of the wave [5].

Scintillation

Scintillation is a flash of light produced in a transparent material by the passage of a particle (electron, ion or high energy photon). Scintillation is a natural molecular property in conjugated and aromatic organic molecules and arises from their electronic structures. Scintillation also occurs in many inorganic materials, including salts, gases and liquids [5].

BER and SNR

The most used characteristics, not only in optical communication, are the bit error rate - BER and the value of the signal to noise ratio - SNR (Signal to Noise Ratio). The BER value depends on the average received optical power, the intensity of scintillation and the noise of the receiver. By selecting suitable dimensions of the receiving aperture, it is possible to increase the intensity of the received optical power and at the same time reduce the negative impact of scintillation. The value of the SNR parameter can be expressed by the relation [5] due to turbulence:

$$SNR = \frac{1}{0{,}31 C_n^2 k^{7/6} l^{11/6}} \tag{2}$$

where C_n^2 is the refractive index structure, $k = 2\pi / \lambda$ is the wave number (an expression suggests that longer wavelengths experience a smaller variance), and l is the link range (m). FSO systems use ON/OFF modulation, for which it is possible to express BER according to equation [3]:

$$BER = \frac{exp(-SNR\ /\ 2)}{(2\pi SNR)^{0.5}} \tag{3}$$

If we assume that the receiving area of the photodetector of the FSO line head is large enough, it is necessary to consider

the influence of the optical beam propagation when calculating the SNR. In this case, the so-called effective SNR (SNR_{eff}), calculated according to [5]:

$$SNR_{eff} = \frac{SNR}{1 + 1{,}33\sigma_i^2 \left[\frac{2l}{k\omega(l)^2}\right]^{5/6}} \quad (4)$$

where σ_i^2 is scintillation index.

B. RF system

Radio frequency (RF) wireless communication systems have existed for many years. These technologies cover a wide range of options tailored to different uses and needs. In RF wireless communication systems, radio waves are used to transmit information between a transmitter (Tx) and a receiver (Rx). RF systems can be classified as terrestrial or space systems. Terrestrial systems include microwave point-to-point, wireless local area networks (WLAN) and cellular mobile radio. Terrestrial microwave systems are limited in distance and the limiting factor is the spread of visibility (LOS).

Radio frequencies or radio waves form part of the electromagnetic spectrum ranging from 3 kHz to 3 THz. The entire RF spectrum is classified into different bands and ranges based on propagation properties. Baseband signals or source signals (eg, audio signals) are in the low frequency range below 30 kHz. This frequency range is classified as very low frequency (VLF), which must be translated into RF before transmission [12].

FSO connections suffer mainly from fog attenuation and are sensitive to allegedly long periods of complete connection loss. When available, the FSO system provides very high bandwidth without the need for licensing. The RF link for FSO / RF hybrid systems should therefore meet the following requirements:

- availability during the period of loss of the FSO link, i.e., during the fog event,
- bandwidth comparable to the FSO link to provide similar performance,
- unlicensed operation.

As it turns out, these requirements are contradictory. High carrier frequencies must be used to achieve high bandwidths. These frequencies, on the other hand, are strongly affected by precipitation and fog, although the attenuation caused by fog is much lower than in the case of FSO connections. Still, due to limited transmission power (EIRP) or limited antenna gain in unlicensed bands and much greater geometric loss for RF connections, the available error margin is also much lower, causing significant availability problems. Especially in times of rain and fog, as is often the case in continental climates, both connections suffer from the weather at the same time. To overcome these problems, it is necessary to increase transmission power, which often requires authorization [12]. On the other hand, lower frequencies are not affected by fog at all. Unlicensed systems in the ISM bands (2.4 or 5 GHz), because they are used for IEEE 802.11a/b/g compliant wireless LAN devices, are only affected by rain or multi-user interference.

C. Hybrid FSO/RF systems

The rapid development of information technology, the overcrowding of the high-frequency spectrum, and the pressure on mobile data services have led to a huge demand for the establishment of high-speed ubiquitous wireless networks. Due to the limitation of data transfer speed, security and power consumption in traditional RF communication, many researchers are looking for new technology to overcome these problems. FSO communication is one of these technologies [13].

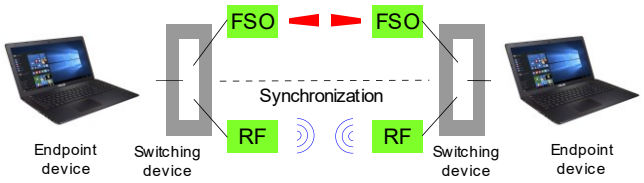


Figure 3 Principle of operation of a hybrid FSO/RF system.

Due to very different carrier frequencies, FSO and RF lines have different channel failures. The short wavelengths used in FSO lines are severely affected by fog scattering and atmospheric scintillation, while high frequency links are particularly sensitive to rain scattering. The above factors make the presence of the FSO link not a conventional technology compared to the method of communicating with optical fibers. On the other hand, RF communication can combat weather conditions. The hybrid structure of the FSO system with a backup RF line combines the advantages of both technologies, such as low cost, unattended applications, high data rates and low latency. The disadvantage is that RF typically provides an order of magnitude less channel capacity compared to optical communications. Other challenges for RF include lack of bandwidth due to licensing, interference sensitivity, and vulnerabilities in detection and eavesdropping. Using both channels cooperatively allow hybrid FSO/RF systems to combine high baud rates of FSO systems and reliability of high frequency connections through a variety channel. As a result, it was possible to create a hybrid network that can eliminate the failures of FSO systems and ensure a network that is almost 100% operational [14].

To improve the performance of the FSO joint and increase its reliability in various weather conditions, the FSO system can be combined with an RF system to form a hybrid FSO/RF system. This hybrid transmission system has two switching configurations [15]:

- Hard Switching,
- Soft Switching

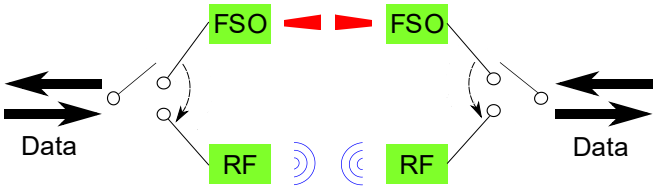


Figure 4 Hard switching scheme of a hybrid FSO/RF system.

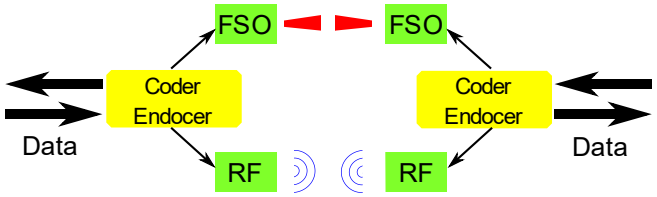


Figure 5 Soft switching scheme of a hybrid FSO/RF system.

Configuration with hard switching, shown in. Figure 4 selects an FSO or RF data transmission channel on both transmitter and receiver side. The communication itself always takes place only on one line. However, the connection is primarily realized by means of an FSO line, the RF channel

serves only as a backup. In this switching configuration, there is feedback between the receiver and the transmitter, allowing the switching of lines on both sides to be coordinated. The main disadvantage of this system is usage of only one line at a time. On the other hand, soft switching utilizes both lines, which is shown in Figure 5. Data is sent to both lines at the same time, leaving only the one that arrives earlier. Thus, if the transmitted data is previously received by the FSO line, the data sent by the RF line is discarded. This method of communication is a bit more complicated, as both lines communicate at the same time. This leads to higher energy consumption [16, 17].

III. Recording Weather Conditions Using a Monitoring Station

The reliability and availability of FSO systems is a complex problem and its solution requires real-time analysis of atmospheric conditions during the entire transmission route. A weather-monitoring station shown in Figure 6 was created for this purpose [18]. The station measures following parameters:

- Visibility.
- Temperature.
- Concentration of particles in the fog.
- Wind speed.
- Humidity.
- Barometric air pressure.

Figure 6 Weather monitoring station together with FSO head and RF line.

IV. Further Research Direction

In my further research, I will work on further research of the transmission channel between FSO heads as well as the effects of the environment on this transmission. As this is a communication based on direct visibility, mirrors for signal reflection could be used to improve communication in the built-up area and thus *ensure* communication even in the event of obstacles on the transmission path. Another important component is the RSSI parameter, i.e., the received signal strength indicator, which determines whether the signal level transmitted by one of the pair of FSO heads is sufficient for communication. I consider the work on hard switching between individual lines of the hybrid FSO / RF system to be a substantial part, as we want to ensure almost 100% functionality of both lines even in bad weather conditions. In the case of hard switching, the RSSI parameter is decisive, and we can decide whether the FSO or the RF line should work.

Acknowledgment

This work was supported by research grant of FEEI TU Košice no. FEI-2021-71.

References

[1] J. Mikolajczyk, Z. Bielecki, et al. “Analysis of free/space optics development,” Metrology and measurement systems, 2017, vol. 24, no 4, pp. 653-674.

[2] M. Uysal, C. Capsoni et al., Optical Wireless Communications An Emerging Technology. Signals and Communication Technology, Springer International Publishing, 2016, 634 p., ISBN: 978-3-319-30201-0.

[3] A. Malik and P. Singh, “Free Space Optics: Current applications and future challenges,” International Journal of Optics, 2015, pp. 1-7, doi: 10.1155/2015/945483.

[4] J. Kaufmann, “Free space optical communications: an overview of applications and technologies,” in Proceedings of the Boston IEEE Communications Society Meeting, 2011.

[5] A. G. Alkholidi and K. S. Altowij, :Free Space Optical Communications - Theory and Practices, a chapter from Wireless Communications. London: IntechOpen, 2014, pp. 159-212, doi: 10.5772/58884.

[6] N. A. M. Nor, Md. R. Islam, et al., “Atmospheric Effects on Free Space Earth-to-Satellite Optical Link in Tropical Climate,” International Journal of Computer Science, Engineering and Applications, 2013, vol. 3, no. 1, pp. 17-36.

[7] M. Yahya, K. Salleh, et al., “Low cost hybrid RF-FSO to reduce rain effect in tropical region,” IEEE International RF and Microwave Conference, Seremban (Malaysia), 2011, pp. 250-253.

[8] A. Viswanath, V. K. Jain, S. Kar, “Performance Evaluation of Satellite-To-Earth FSO link in Presence of Turbulence and Weather Conditions for different IM schemes,” Twenty Second National Conference on Communications, Guwahati (India) 2016, pp. 1-5.

[9] M. Esmail, H. Fathallah, M. Alouini, “Outdoor FSO Communications Under Fog: Attenuation Modeling and Performance Evaluation,” IEEE Photonics Journal, 2016, vol. 8, no. 4, pp. 1-22.

[10] H. Singh, D. P. Checchi, “Performance Evaluation of Free Space Optical (FSO) Communication Link: Effects of Rain, Snow and Fog,” International Conference on Signal Processing and Integrated Networks, Noida (India), 2019, pp. 387-390.

[11] J. Haby, "Wet vs. dry snow," [Online]. Available: http://theweather-prediction.com/ habyhints2/650/.

[12] O. Ugweje, “Radio Frequency and Wireless Communications Security;” 2004, doi: 10.1002/047148296X.tie151.

[13] W. A. Alathwary, E. S. Altubaishi, et al., “On the Performance Analysis of Decode-and-Forward Multi-Hop Hybrid FSO/RF Systems With Hard-Switching Configuration,” IEEE Photonics Journal, 2019, vol. 11, no. 6,pp. 1-12, doi: 10.1109/JPHOT.2019.2949859.

[14] K. Nock, C. Font, M. Rupar, “Adaptive Transmission Algorithms for a Hard-Switched FSO/RF Link,” Military Communications Conference, Baltimore (USA), 2016, pp. 877-881.

[15] I. E. Lee, Z. Ghassemlooy, et al., “Practical Implementation and Performance Study of a Hard-Switched Hybrid FSO/RF Link under Controlled Fog Environment,” 9th International Symposium on Communication Systems, Networks & Digital Sign, Manchester (UK), 2014, pp. 368-373.

[16] W. M. R. Shakir, “Performance evaluation of a selection combining scheme for the hybrid FSO/RF system,” IEEE Photon. J., 2018, vol. 10, no. 1, pp. 1-10, doi: 10.1109/JPHOT.2017.2771411

[17] W. Zhang, S. Hranilovic and C. Shi, “Soft-Switching Hybrid FSO/RF Links Using Short-Length Raptor Codes: Design and Implementation, “IEEE Journal on Selected Areas in Communications, December 2009, vol. 27, no. 9, pp. 1698-1708. ISSN 0733-8716. DOI: 10.1109/JSAC.2009.091219

[18] M. Lapčák, “Design of hard switching for a hybrid FSO/RF transmission system based on the prediction of the received signal strength parameter,” Diploma thesis, Technical university of Košice, p. 74, 2020.

Dielectric nanofluids facing sustainable development

[1]*Miloš ŠÁRPATAKY (1st year)*
Supervisor: [2]Juraj KURIMSKÝ

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

[1]milos.sarpataky@tuke.sk, [2]juraj.kurimsky@tuke.sk

***Abstract*— Functioning of high voltage equipment depends on good heat transfer and insulation qualities of combined solid and liquid insulation. Research of fluid insulation is crucial for this insulation pair as the heat transfer medium. Striving to create the best possible insulating liquid with environmentally friendly properties is the main goal of this field of study. There is a need to have enough comparable information to prove that biodegradable insulation fluids or nanofluids meet or overcome qualities of the oils currently in use. This paper compares different insulating liquids, its properties and describes the possible improvement of used fluids by nanoparticles to get better and more efficient insulating and heat transfer medium.**

***Keywords*— Biodegradable oils, esters, insulation properties, nanofluid, transformer insulation.**

I. Introduction

Power transformers are of crucial importance devices in power engineering. They are used for transformation of voltage from one level to another, or for galvanic separation of different electrical systems on the same level of voltage. To ensure the functioning of this device, the heat transfer and electrical insulation need to be secured. Until today, the most frequently used insulation in power transformers are oils. Their important properties are viscosity, thermal conduction, fire point, biodegradability as physical properties and breakdown voltage, dissipation factor and resistivity of fluid as dielectric properties [22]. A huge number of transformers are filled with mineral oil. However, they are currently considered unsatisfactory due to their environmental footprint. The effort of today's research is to replace mineral oils with a more environmentally friendly alternative e.g. natural (vegetable oils) or synthetic esters [30]. Natural or synthetic esters as environmental friendly oils have biodegradability over 80 percent whereas mineral oils reach values less than 30 percent [33]. Dielectric properties of mineral oils are comparable or better than in esters what is one of the reasons for the usage of mineral oils. Current research has shown the possibility of improving dielectric and thermal properties by modifying liquids with nanoparticles. This mixture of base fluid and nanoparticles is called nanofluid. Nanofluids are considered as a future substitute for the current generation of insulation and heat transfer fluids used in high voltage equipment [19]. The mixture of nanoparticles and biodegradable base fluids seems to be a good solution for maintaining sustainable development.

II. Insulating fluid characterization

To compare different insulating fluids, the quantities are given that describe its heat transfer and dielectric properties or other physical properties. Physical properties are characterized by viscosity, flash point, flash point, freezing point, and thermal conductivity. Electrical properties are described by AC breakdown voltage, dissipation factor and resistivity. Chemical properties are related to reactions in a fluid or with other materials in high voltage equipment. An important chemical factor that affects the main insulation quantity is water content of the liquid [19], [26]. Next part of this paper deals with a description of properties important for insulation fluids, for more information and insulation fluid requirements see [5].

A. Physical properties

Viscosity is in general resistance to flow that means we need value as low as possible. In other words, it can be also described as an ability of the liquid to transfer heat by conduction. Cooling of high voltage system by conduction is major heat removal system. With high fluid viscosity, there may appear hot spots that are caused by inefficient circulation [21], [22].

Flash and fire points are important values to describe the safety of insulating fluids in extreme conditions at particular high temperature. Flash point shows the tendency of fluid to create a flammable compound with the atmosphere. This value designates very combustible materials [19]. The fire point is a safety value connected with the ignition of material [2].

Pour point is the low value of temperature down to approximately -65 °C when liquid is no more flowable. This value is important for example when the power transformer is operated in the outer environment in geographical regions with very cold seasons [33].

Thermal conductivity describes fluids ability to transfer heat and also shows its temperature dependence.

B. Chemical properties

Acidity is value describing the chemical stability of fluid. It is measuring of number of acidic components in the fluid. Not aged or unused fluid should have no acidic substances, however working device after some time creates acidic derivates as the product of oxidation [19], [33].

Moisture in insulating fluids is very dangerous for the high voltage equipment. Moisture in every of its state influences mainly electrical properties of a fluid. It is very important in practical usage, but also in experimental conditions to have a low amount of moisture or to lower moisture in fluid before its

usage in a power transformer or for experimental measurements [7].

C. Electrical properties

AC breakdown voltage is the most important insulating property as it describes electrical strength of the insulating fluid or in other words, it describes the amount of electrical stress that fluid can withstand without collapsing. Breakdown voltage is a value when current start to flow between electrodes so fluid (or solid insulation) lose its insulating properties. This value determines the suitability of insulation for application in high voltage equipment [10], [14].

Dissipation factor is mainly connected with dielectric loss caused by impurities or oxidation derivates and also very important value that increases dissipation factor is content of moisture in liquid. This factor is one of the options to compare the quality of insulation materials [7], [19].

Reference [7] describes **resistivity** of material as **"***ratio between the intensity of a continuous electric field and the steady state value of the current density in the material*". Higher value of resistivity means better insulating properties as it was with breakdown voltage.

TABLE I
EXAMPLE OF REQUIREMENTS FOR INSULATING ESTER FLUID [26]

Properties	Unit	Standard for evaluation	Required value
Density at 20 °C	Kg/dm^3	ISO 3675	Max. 1
Viscosity at 40 °C	mm^2/s	ISO 3104	Max. 35
Flash point	°C	ISO 2719	250
Fire point	°C	ISO 2592	300
Pour point	°C	ISO 3016	Max. -45
AC breakdown voltage	kV	IEC 60156	Min. 45
Dissipation factor at 90 °C	-	IEC 60247	Max. 0,03
Resistivity at 90 °C	GΩm	IEC 60247	Min. 2

Table 1 shows requirements specified in IEC 61099 for fresh synthetic ester, but for other fluids, it can be different according to basic features of insulating fluids, for example, mineral oils are not able to reach as high fire point as synthetic ester [26].

III. SHORT HISTORICAL REVIEW ON INSULATING FLUIDS

In reference [19] a nice historical review can be found. Accordingly, short history information is compiled: First oil-immersed transformer was developed in 1890 with petroleum-based liquid. Paraffin oil with small viscosity and outstanding insulation properties was used until 1925 and then it was replaced because of its high pour point and insoluble particles caused by oxidation that reduced its heat transfer and lifespan. For short time, paraffin oils have been replaced by naphthenic oils, which have provided a higher pour point as well as better oxidative stability. However, with high flammability and with the fact it was environmental contaminant it had to be replaced, too. For 40 years, till 1970, the most commonly used were PCBs oils (polychlorinated biphenyls). They were found toxic, so they were no longer environmentally acceptable. Finally, PCBs were banned in 1978[19], [20].

After PCB oils were banned mineral oils and silicone fluids were introduced, and its usage continues till nowadays. The demand for better properties of insulating fluids, mainly the importance of the environmentally friendly features, like biodegradability, led to the need for ecologically safe fluids, like natural and synthetic esters. Biodegradability of natural esters is around 98% and synthetic esters around 80%, while mineral oils have biodegradability under 30% and silicone fluids are almost undegradable. Research of nanofluids started in the 1990s and P. Kopčanský et al. started to study nanofluids in the publication "Magneto-dielectric behavior of concentrated magnetic fluid" in 1996 and from that point, there were many papers written in Slovakia which dealt with nanofluids [12]. Technical University of Košice (TUKE) is in fourth place when we talk about publications about nanofluids worldwide, after NECEPU, XJTU and Chongquind University all from the Republic of China [20]. Cooperation of TUKE and SAV (Slovak Academy of Sciences) from Slovakia made numbers of publications for example [9], [13], [15] or [23] where the greatest merit for the development of this area of research had P. Kopčanský, K. Marton, F. Herchl and M. Timko.

Till today, the most frequently used oils worldwide are mineral oils (MO), silicone fluids (SF), natural esters (NE) and finally synthetic esters (SE) [30]. Difference between these oils is in various factors. They are discussed in following chapter IV.

IV INSULATING LIQUIDS COMPARED

A. Mineral oils

Mineral oils (MO) are the most frequently used oils worldwide. It is mainly because of their excellent performance in practice. Generally, MO is made of fossil fuels (crude oil) that contain hydrocarbon compounds of different bonds [4], [33]. The critical issue of this kind of oil nowadays is its low biodegradability and the fact that it is a potential environmental contaminator. Safety for people and environment is endangered by its relatively low flash and fire point, when it could cause great damage after ignition (1000 kg of burnt, under unfavorable conditions, MO can emit 10 kg of harmful substances to the atmosphere caused by a short circuit in power transformer [27]) and leakage of power transformer oil to the outer environment can result in a serious ecological problem of the affected region [11], [27]. However, Table 2 shows excellent values of some important properties when we talk about heat transfer and insulation. For example, the viscosity of MO is significantly better in comparison with other oils. The most expressive value when we talk about electrical properties is dissipation factor of MO. Dissipation factor relates to dielectric loss, so MO have outstanding efficiency, low dielectric loss and therefore also little financial loses if we talk about operation of a power transformer.

B. Silicone fluids

The evident difference between the silicone fluids (SF) and MO is their flammability. SF is less-flammable liquid with comparable electrical properties with MO. In Table 2, there is a difference in fire point between SF and MO from 145 to 240 °C what makes a significant contrast in safety between them. SF is chemically known as polydimethylsiloxane and this insulation liquid is barely used worldwide mainly because of a

higher price than MO, and because of ecological issues connected to the biodegradability of this fluid [29], [33].

C. Synthetic esters

A major component of synthetic esters (SE), which are used for power transformer electric insulation, is pentaerythritol tetra ester and this fluid is developed from chemicals. However, by reason of its chemical structure, SE is biodegradable. The first transformer filled with this insulation liquid was introduced in 1976. Nowadays, it is applied mainly in distribution and special transformers at relatively low levels of voltage or power ranges [19], [26]. From Table 2, is clear that properties of SE are comparable to MO or they are even better. The noticeable disadvantage of this fluids is dissipation factor and low commercial availability that relates to economic factors [22].

D. Natural esters

Natural esters (NE) are the most "eco-friendly" solution for insulation and heat transfer in high voltage equipment. Availability of NE is growing year by year in insulation fluid market that makes it the most likely option of refilling the power transformers in future. NE is refined vegetable oil or in other words, it is a plant-based ester derived from crops. These oils are mainly extracted from the seed, but also other parts of the plant can yield oil [19]. The most significant advantage of these oils is biodegradability as it was mentioned. This fact is connected with the safety of this insulation liquid during possible leakage, but its high fire and flame point secure its reliability what confirms that in 2014 there were over 600 000 transformers filled with NE worldwide with no reported fire or environmental incident [24]. Disadvantages of this insulating fluid are high pour point and not excellent dissipation factor. However, the electrical properties of dielectric fluids can be improved by adding nanoparticles. NE may be an ideal fluid for refinement because of its biodegradability. Moreover, improvement of NE with nanoparticles may create "eco-friendly" insulation fluid with enhanced properties.

A comprehensive compressed review of the current insulating fluids and their physical and electric properties are in Table 2.

TABLE II
DIFFERENT PROPERTIES OF BASIC INSULATING FLUIDS

Properties	MO	SF	NE	SE	Reference
Biodegradability in 21 days [%]	< 30	very low	97-99	80-89	[19], [27], [33]
Viscosity at 40 °C [mm²/s]	3-16	35-40	16-37	14-30	[8], [22]
Flash point [°C]	100-175	300-310	250-316	260-343	[22], [27], [33]
Fire point [°C]	110-185	330-350	300-370	300-322	[22], [27], [33]
Pour point [°C]	-30 to -63	-50 to -60	-10 to -33	-40 to -60	[22], [27], [33]
AC breakdown voltage [kV]	30-85	35-60	49-97	43-75	[19], [22], [27], [33]
Dissipation factor at 90 °C [-]	<0,001	0,0016	0,005-0,009	0,001-0,03	[33]
Resistivity at 90 °C [GΩm]	10^{13}-10^{15}	10^{14}	10^{13}-10^{14}	10^{13}	[19], [33]

IV. NANOFLUIDS

Enhancement of properties in insulation fluid is possible by nanoparticles (NP). There are used three main groups of nanoparticles, namely insulating (Al_2O_3, SiO_2 or SiO), conductive (Fe_3O_4 or ZnO) and semi-conductive (TiO_2, Fe_2O_3, CuO or CuO_2). The typical diameter of nanoparticle in the insulating fluid is in the range of 10 to 60 nm [17]. Nanofluid is created by one-step or two-step process. One-step process involves contemporaneous synthesis and dispersion of nanoparticles into the base fluid. Two-step processes have two phases preparation of nanoparticles which are usually in the form of powder and the second phase is a dispersion of this prepared nanoparticles into fluid during magnetic stirring or ultrasonication. Addition of dispersant to nanofluid during this process enhances the stability of nanofluid [1]. Simple illustration of this process is in Fig. 1.

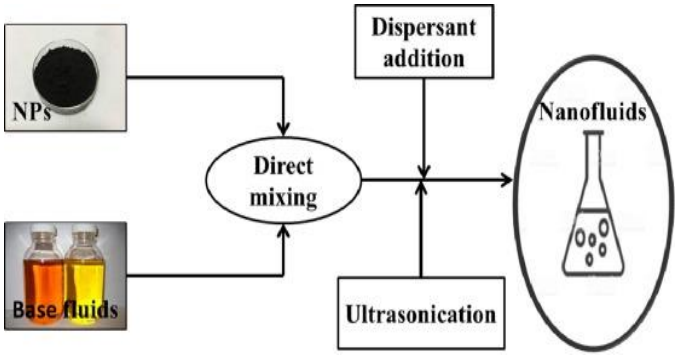


Fig. 1. Preparation of nanofluid [21].

Origin of the word "nanofluid" is associated with Choi et al. when they did research about connecting fluids and nanoparticles back in 1995 [19]. From that point, a lot of research has been made and number of scientific papers about this topic rapidly increased. Till 2005 publication on this topic did not exceed number 100 per year. From this year they started to grow and in 2019 there were over 3000 papers published, so that fact confirms actuality and future direction of this field of study [6].

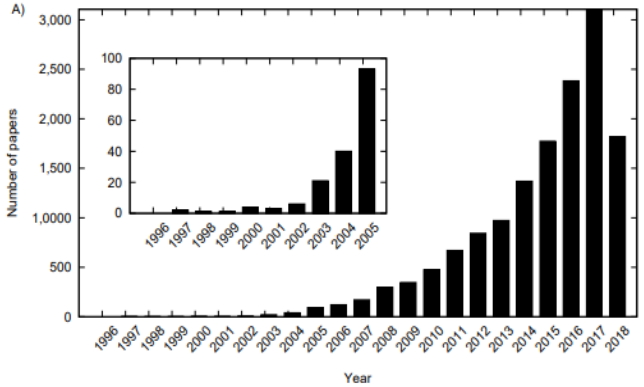


Fig. 2. Number of papers according to Web of Science with the word "nanofluid" in the title [6].

Moreover, in the year 2019, there were over 3600 publications in 2020 over 4100 and in the current year, there are already 283 at the beginning of February.

TABLE III
BREAKDOWN VOLTAGE (BV) IN DIFFERENT NANOFLUIDS

Base fluid	NP	NP size [nm]	Concentration [%]	BV [kV]	Reference
NE	Al_2O_3	50	0,004	63,8	[14]
NE	SiC	50	0,004	71,3	[14]
NE	-	-	-	61,3	[14]
MO	Fe_3O_4	10	0,005	52,68	[6]
MO	Fe_3O_4	10	0,01	54,41	[6]
MO	Fe_3O_4	10	0,015	52,38	[6]
MO	TiO_2	<20	0,05 [g/L]	53,25	[3]
MO	TiO_2	10	0,06 [g/L]	51,24	[28]
NE	C60	4-6	0,1 [g/L]	68,2	[31]
NE	C60	4-6	0,15 [g/L]	64,1	[31]
NE	C60	4-6	0,2 [g/L]	58,8	[31]

Table 3 indicates the behaviour of different nanofluids according to base fluid, type of NP, size of NP, its concentration and breakdown voltage. Same base fluid with different nanoparticles shows different results. NE with silicon carbide has a higher breakdown voltage than the same fluid but with aluminium oxide. However, both of these nanofluids express improvement in comparison with clear base fluid. Another factor that affects the results of breakdown voltage measurements is a concentration of nanoparticles in the base fluid. Same fluid with same NP of the same size, but with different concentration present different values. In every combination of the base fluid and NP, there is a value of concentration that shows better results than others. In this example, the best option of concentration of magnetite NP is 0,01 % of NP in the fluid. If we add or reduce amount of NP in fluid breakdown voltage will decrease that is approved by C60 (Fullerene) nanoparticles in vegetable oil [31]. The last factor that will be mentioned is the size of NP. MO with titanium dioxide NP with almost the same concentration have different results. In Table 3 can be seen that nanoparticles with lower size enhance insulation properties of base fluid more significantly. Moreover, other factors can affect the results of various investigations like different base fluid with the same type, size and concentration of NP will show distinct results. This fact is important in future practice because it will be necessary to distinguish different types of nanoparticles for different types of insulating fluids. Other factors that may affect measurements are content of moisture, the process of production of nanofluid, aging, temperature of fluid, pressure etc. [21].

V. Present and Future Direction of Insulating Fluids

Mineral oils are the most frequently used insulating fluid worldwide as was mentioned. The lifespan of mineral oil specimen according to Mohammad R. et al. at 60 °C is around 28,5 years [16]. In real conditions aging of insulating oil is influenced by various factors like temperature, working conditions, quality of oil and so on. After oils lifespan, there are few options to solve this problem concretely retro filling transformer oil or improvement of its qualities by reconditioning. Process of reconditioning is connected mainly with vacuum processing and filtering of aged liquid. High Vacuum Oil Purifier and filters are used to remove contaminants from transformer oil mainly moisture, acids and sludges created by the oxidation process. With 2 cycles of reconditioning values of breakdown voltage test in MO can be almost doubled in experimental conditions [25]. This process of reconditioning is one of the options but is more suitable for maintenance of used oil because the ability for reinstatement decreases with the aging of insulation oil and its quality. Sooner or later transformer oil must be retro filled. The future of this field depends on which oil will be mostly used in high voltage equipment.

A. Retro filling of transformer oil

Retro filling of transformer oil means that insulation fluid in the transformer is exchanged by another same or different fluid. This process begins with pumping of old oil out of the transformer tank than windings and core are rinsed with heated new fluid, next step is fully drained of all fluids out of the transformer and finally refilling of new heated fluid into power transformer [36]. After this process, there are 3% of old transformer oil contained in the refilled one [35]. A good example of retro filling is an exchange of mineral transformer oil by synthetic ester in Berlin (Germany) where 350 power transformers were retro filled in the 1990s led by European utility Vattenfall [32]. After years of monitoring synthetic ester in practice this insulation oil shows these advantages:

- Prolonging of oil (also transformer) lifespan.
- Esters have much better moisture saturation, so they are not as affected by moisture content as MO.
- They are safer than MO, that means they have higher fire and flash point, and they are readily biodegradable, unlike MO which is toxic and barely biodegradable.
- Financial savings connected with lowering of operating expenses, insurance cost (NE are much safer), prolonged lifespan, removal of possible environmental damage [26], [34].

However, there are possible technical complications connected with application of esters in high voltage equipment. The price of esters is dependent on feedstock cost, storage and management are problematic and challenging, the different base of fluids can show different quality, approval of this kind of insulation fluid by manufacturers can be an additional problem, pour point must be modified by additional depressants [19].

B. Nanofluids as the future option of transformer oil

Power transformers need to have oil with excellent cooling and insulation abilities that can be enhanced by nanoparticles. For example, MO improved by conductive nanoparticles is able to reduce hot spot and top-oil temperature in power transformer by almost 5 °C if we compare it to base fluid (MO) [18]. Reliability and lifespan of the existing transformer can be

prolonged and improved by nanoparticles that in the future might be added or whole insulation fluid may be replaced by nanofluid to enhance the quality of insulation. As the possible future alternative of insulation fluid nanofluids have advantages and disadvantages. Let us start with advantages:

- Nanofluids have better AC and impulse breakdown voltage as compared to the base fluid.
- AC breakdown strength of nanofluids is less affected by the content of moisture than MO so the lifespan of the transformer is prolonged.
- Partial discharge characteristics are improved.
- transformer oil-based nanofluids are less influenced by aging that relates to lifetime and operational reliability.
- Nanoparticles enhance the thermal conductivity of the base fluid and show better cooling performance [18].

On the other hand, there are disadvantages connected with nanofluids which slow down the development and use of nanofluids in practice:

- Stability of nanofluids is theoretically confirmed when the size of nanoparticles is not higher than 100 nm, but the challenge is to eliminate forces like gravity because of sedimentation and agglomeration of nanoparticles in a fluid.
- Possible health issues with nanofluids mainly connected with nanoparticles can cause various health problem after reaching the lungs, digestive or nervous system.
- Some kinds of nanoparticles may be toxic that cause problem with mentioned health issues for example carbon and graphite nanoparticles can cause skin and respiratory infections.
- Environmental problems can be caused by toxic nanoparticles spreading in the air and reaching water, vegetation, or soil.
- High manufacturing cost inhibits the implementation of nanofluids into practice [18].

VI. CONCLUSION: THE VISION FOR FURTHER RESEARCH

To sum up, the change of insulation fluids in future is almost certain. Our vision for further research raised from the biggest challenge of today: an implementation of environmentally friendly nanofluids, biodegradable fluids, or their mixture into the practice. The future research will be focused on the research and confirmation of all good features that nanofluids and biodegradable oils can offer. Before the implementation, there is a need to approve all insulating oils properties necessary for the functioning of a transformer by many measurements and experiment in laboratory conditions and then confirming the results in practice.

FUNDING SOURCES

The research work of nanofluid dielectrics has been funded mainly by:

Scientific Grant Agency Ministry of Education, Science, Research and Sport of the Slovak Republic and SAS (VEGA) VEGA 2/0011/20, VEGA 1/0154/21.

The Slovak Research and Development AgencyAPVV-15-0438, APVV-18-0160.

ACKNOWLEDGMENT

This publication is the monitored impact of the EU ERDF, Operational Program Integrated Infrastructure 2014-2020 project implementation: Innovative Testing Procedures for 21st Century Industry ITMS: 313011T565

REFERENCES

[1] AHMAD, F. et al. State-of-Art in Nano-Based Dielectric Oil: A Review. In *IEEE Access* . 2019. Vol. 7, s. 13396–13410. .

[2] AZIS, N. et al. Suitability of Palm Based Oil as Dielectric Insulating Fluid in Transformers. In *Journal of Electrical Engineering and Technology* . 2014. Vol. 9, no. 2, s. 662–669. .

[3] DU YUE-FAN et al. Breakdown properties of transformer oil-based TiO2 nanofluid. In *2010 Annual Report Conference on Electrical Insulation and Dielectic Phenomena* [online]. West Lafayette, IN: IEEE, 2010. s. 1–4. [cit. 2021-02-04]. Dostupné na internete: <http://ieeexplore.ieee.org/document/5724082/>.

[4] EBERHARDT, R. et al. Partial discharge behaviour of an alternative insulating liquid compared to mineral oil. In *2010 IEEE International Symposium on Electrical Insulation* [online]. San Diego, CA, USA: IEEE, 2010. s. 1–4. [cit. 2021-02-03]. Dostupné na internete: <http://ieeexplore.ieee.org/document/5549776/>.

[5] EKLUND, M. Mineral insulating oils; functional requirements, specifications and production. In *Conference Record of the 2006 IEEE International Symposium on Electrical Insulation* . 2006. s. 68–72. .

[6] FAL, J. et al. Nanofluids in the Service of High Voltage Transformers: Breakdown Properties of Transformer Oils with Nanoparticles, a Review. In *Energies* . 2018. Vol. 11, no. 11, s. 2942. .

[7] FERNÁNDEZ, I. et al. Effect of TiO2 and ZnO Nanoparticles on the Performance of Dielectric Nanofluids Based on Vegetable Esters During Their Aging. In *Nanomaterials* . 2020. Vol. 10, no. 4, s. 692. .

[8] GNANASEKARAN, D. - CHAVIDI, V.P. Vegetable Oil: An Eco-friendly Liquid Insulator. In GNANASEKARAN, D. - CHAVIDI, V.P. *Vegetable Oil based Bio-lubricants and Transformer Fluids* [online]. Singapore: Springer Singapore, 2018. s. 101–124. [cit. 2020-11-02]. ISBN 978-981-10-4869-2Dostupné na internete: <http://link.springer.com/10.1007/978-981-10-4870-8_6>.

[9] HERCHL, F. et al. Breakdown and partial discharges in magnetic liquids. In *Journal of Physics: Condensed Matter* . 2008. Vol. 20, no. 20, s. 204110. .

[10] HUANG, Z. et al. Significantly enhanced electrical performances of eco-friendly dielectric liquids for harsh conditions with fullerene. In *Nanomaterials* . 2019. Vol. 9, no. 7. .

[11] KARMAKER, A.K. et al. Investigation and Analysis of Electro-physical Properties of Biodegradable Vegetable Oil for Insulation and Cooling Application in Transformers. In *Journal of Electronic Materials* . 2020. Vol. 49, no. 1, s. 787–797. .

[12] KOPČANSKÝ, P. et al. Magneto-dielectric behaviour of concentrated magnetic fluid. In *Journal of Magnetism and Magnetic Materials* . 1996. Vol. 157–158, s. 587–588. .

[13] KOPČANSKÝ, P. et al. The DC dielectric breakdown strength of magnetic fluids based on transformer oil. In *Journal of Magnetism and Magnetic Materials* . 2005. Vol. 289, s. 415–418. .

[14] KOUTRAS, K.N. et al. The Influence of Nanoparticles' Conductivity and Charging on Dielectric Properties of Ester Oil Based Nanofluid. In *Energies* . 2020. Vol. 13, no. 24, s. 6540. .

[15] KUDELCIK, J. et al. Dielectric breakdown in mineral oil ITO 100 based magnetic fluid. In *Physics Procedia* . 2010. Vol. 9, s. 78–81. .

[16] MESHKATODD, M.R. Aging Study and Lifetime Estimation of Transformer Mineral Oil. In *American Journal of Engineering and Applied Sciences* . 2008. Vol. 1, no. 4, s. 384–388. .

[17] PRIMO, V.A. et al. Improvement of transformer liquid insulation using nanodielectric fluids: A review. In *IEEE Electrical Insulation Magazine* . 2018. Vol. 34, no. 3, s. 13–26. .

[18] RAFIQ, M. et al. A Review on Properties, Opportunities, and Challenges of Transformer Oil-Based Nanofluids. In *Journal of Nanomaterials* . 2016. Vol. 2016, s. 1–23. .

[19] RAFIQ, M. et al. Sustainable, Renewable and Environmental-Friendly Insulation Systems for High Voltages Applications. In *Molecules* . 2020. Vol. 25, no. 17, s. 3901. .
[20] RAFIQ, M. et al. The impacts of nanotechnology on the improvement of liquid insulation of transformers: Emerging trends and challenges. In *Journal of Molecular Liquids* . 2020. Vol. 302, s. 112482. .
[21] RAFIQ, M. et al. Transformer oil-based nanofluid: The application of nanomaterials on thermal, electrical and physicochemical properties of liquid insulation-A review. In *Ain Shams Engineering Journal* [online]. 2020. [cit. 2020-11-02]. . Dostupné na internete: <http://www.sciencedirect.com/science/article/pii/S2090447920301805>.
[22] RAFIQ, M. et al. Use of vegetable oils as transformer oils – a review. In *Renewable and Sustainable Energy Reviews* . 2015. Vol. 52, s. 308–324. .
[23] RAJNAK, M. et al. Dielectric response of transformer oil based ferrofluid in low frequency range. In *Journal of Applied Physics* . 2013. Vol. 114, no. 3, s. 034313. .
[24] RAPP, K.J. et al. Application of Natural Ester Insulating Liquids in Power Transformers. In . 2014. s. 7. .
[25] RODIAH, Y. et al. Life Expectancy of Transformer Insulation System by Reconditioning. In *2018 4th International Conference on Science and Technology (ICST)* . 2018. s. 1–4. .
[26] ROZGA, P. et al. A Review on Synthetic Ester Liquids for Transformer Applications. In *Energies* . 2020. Vol. 13, no. 23, s. 6429. .
[27] ROZGA, P. PROPERTIES OF NEW ENVIRONMENTALLY FRIENDLY BIODEGRADABLE INSULATING FLUIDS FOR POWER TRANSFORMERS. In . 2013. s. 7. .
[28] SIMA, W. et al. Effects of conductivity and permittivity of nanoparticle on transformer oil insulation performance: experiment and theory. In *IEEE Transactions on Dielectrics and Electrical Insulation* . 2015. Vol. 22, no. 1, s. 380–390. .
[29] STYLIANOS, K. DISTRIBUTION TRANSFORMERS AND MAINTENANCE. In . s. 58. .
[30] TOKUNAGA, J. et al. Palm fatty acid ester as biodegradable dielectric fluid in transformers: a review. In *IEEE Electrical Insulation Magazine* . 2019. Vol. 35, no. 2, s. 34–46. .
[31] YAO, W. et al. Preparation and dielectric properties of natural ester based insulating nano-oil modified by fullerene filler. In *2017 IEEE 19th International Conference on Dielectric Liquids (ICDL)* [online]. [s.l.]: IEEE, 2017. Dostupné na internete: <https://dx.doi.org/10.1109/icdl.2017.8124723>.
[32] A New Lease on Transformer Life: Reasons to Retrofill. In *Electric Energy Online* [online]. [cit. 2021-02-08]. Dostupné na internete: <http://electricenergyonline.com/energy/magazine/1225/article/A-New-Lease-on-Transformer-Life-Reasons-to-Retrofill.htm>.
[33] Comparative evaluation of alternative fluids for power transformers. In *Electric Power Systems Research* . 2013. Vol. 98, s. 58–69. .
[34] Retrofilling. In *MIDEL* [online]. [cit. 2021-02-04]. Dostupné na internete: <https://www.midel.com/midel-in-use/retrofilling/>.
[35] Retrofilling for environmental protection, Berlin. In *MIDEL* [online]. [cit. 2021-02-05]. Dostupné na internete: <https://www.midel.com/case-studies/retrofilling-midel-increases-environmental-protection/>.
[36] What are the requirements when retrofilling transformers with Midel fluids? In *Insulect* [online]. [cit. 2021-02-05]. Dostupné na internete: <https://insulect.com/energy-blog/what-are-the-requirements-when-retrofilling-transformers-with-midel-fluids>.

Diagnostics methods of insulators in wet and polluted conditions

[1]*Ľuboš ŠÁRPATAKY (1st year)*
Supervisor: [2]Bystrík DOLNÍK

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

[1]lubos.sarpataky@tuke.sk, [2]bystrik.dolnik@tuke.sk

***Abstract*—Transmission and distribution of electricity are a significant part of the electricity supply system. A decrease in the reliability of the electricity supply can affect our daily life. The part of this system that provides insulation of power lines is the insulator. Insulators protect the transmission and distribution lines from electric connection with transmission and distribution towers, which hold individual wires above the ground. If the insulator is damaged by the influence of any stress, a failure like a short circuit can occur. Two of them are stresses caused by pollution that covers the surface of the insulator and the level of humidity that can affect withstand voltage of the insulator. Therefore, it is important to find ways how to monitor and prevent insulators against a decrease in electric strength caused by stresses. This paper describes some methods by which insulators can be monitored and can help to improve failure prediction in polluted and wet conditions.**

***Keywords*—Insulator, pollution, flashover, leakage current, numerical models**

I. INTRODUCTION

Insulators are one of the basic components which are used on transmission and distribution lines. The main function is to isolate the phase conductor from other phases and insulate the conductor from the ground.

Insulators are mostly at risk when they are exposed to various stresses. The first is electrical stress which affects insulator constantly due to electric and magnetic field and in situations as short circuit or overvoltage.

The second stress that influences the performance of the insulator is the weather. There are different weather conditions in different environments, and it is important to choose the insulator adapted to these conditions.

The next is pollution which can cause a reduction of flashover voltage of the insulator by tens of percent. Pollution is defined by equivalent salt deposit density (ESDD), which represent different environments. The most dangerous areas are that with heavy industry and coastal areas.

II. INSULATORS

A. Classification of Insulators

Generally, insulators are assembled from 2 parts. Metal fittings are normally made from iron, steel, or aluminum and insulating materials where basic materials are porcelain, glass, and polymeric materials. Two types of insulators are mostly used to suspend conductors from towers on transmission lines. Disk insulators which have high tensile strength and long-rod insulators where voltage is distributed more evenly than on disk insulators [1].

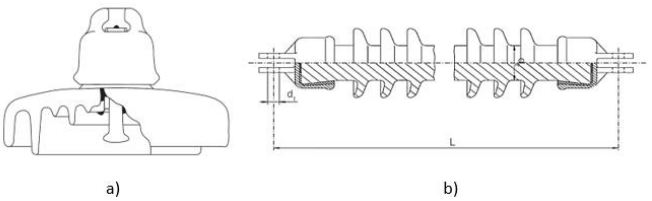


Fig. 1 a) disk insulator, b) long-rod insulator [1].

B. Stresses on insulators

Two stresses influence the performance of insulators the most. Pollution and humidity. In general, when the Obennaus pollution flashover model is used, a surface conductivity is affected by the amount of soluble salt contained in the contamination layer and the flashover voltage of the polluted insulator is related to this surface conductivity. When the wetting process of the contamination layer on the insulator surface starts, the conductivity of the insulator surface increase. To determine the amount of soluble salts in the contamination layer, equivalent salt deposit density (ESDD) is used, where values $\leq 0{,}06$ mg/cm^2 of ESDD are light pollution level and values $> 0{,}1$ mg/cm^2 are heavy pollution level. The contamination level has non-soluble deposit density (NSDD) which forms a sealant layer for the ESDD. The liquid electrolyte that is created on the surface of the insulator can be measured as surface conductivity (μS) [2], [3], [4].

III. FLASHOVER PERFORMANCE OF INSULATORS

Failures of high voltage insulators on transmission lines can lead to transmission line outages, thereby every failure of the insulator can decrease system reliability. One form of insulator failure is flashover which is unintended disruptive electric discharge over or around the insulator [5].

K. A. Aravind, P. Rajamani, B. Krishna made a basic measurement on composite and porcelain insulator where they apply AC breakdown voltage, 50 % impulse flashover voltage of standard lightning, and switching wave with positive and negative polarity.

They find out that the AC breakdown voltage is higher for composite post insulators when compared to porcelain insulators. 50 % lightning impulse flashover voltages are almost the same for both porcelain and composite post insulators. 50 % switching impulse flashover voltage is high for composite post insulators compared to porcelain post insulators. The performance of composite insulators is better

than porcelain post insulators for a longer duration of voltages which are switching impulse voltage and AC voltage [6].

IV. Leakage Current Measurement

A. Leakage current

Leakage current flows on the surface of the insulation when the insulation is subjected to an electric field. If the insulation is not a perfect insulator, there is always conducting layer on the surface of the insulator. This conductive layer on the surface of the insulator is formed because of comprehensive reflection of factors such as pollution, humidity, voltage, and other factors on the surface.

The conductive layer that is formed on the surface is caused by contamination of the insulator that decreases the value of resistance. Reduction of resistance caused an increase of leakage current [3], [7], [8].

B. Measurement methods

Leakage current could be measured in different ways with different equipment.

J. Dadashizadeh Samakosh and M. Mirzaie analyze leakage current during the aging process of SiR insulator. Variable shunt resistors (R_{sh}) are used as leakage current measuring device. They also use a high voltage probe (PINTEK HVP-40) with the ratio of 1000 to 1 and maximum AC voltage of 28 kV for measuring of applied voltage waveform depicts a two-channel digital oscilloscope (RIGOL DS1102E) to record the applied voltage and LC waveforms as you can see in fig. 2 [9].

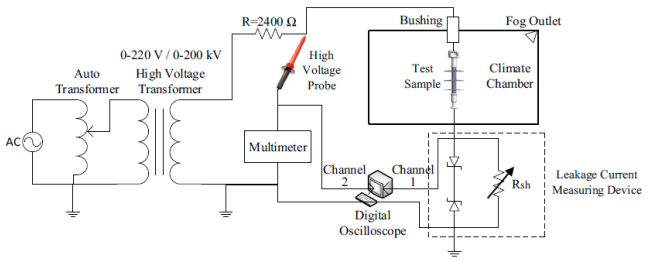


Fig. 2 Schematic diagram of the experimental set-up, [9].

Isaias Ramirez, Ramiro Hernández, and Gerardo Montoya presents an analysis of the monitoring of various types of insulators, namely, porcelain, room temperature vulcanizing silicone (RTV)-coated porcelain, toughened glass, and nonceramic insulators. The monitoring system classifies leakage current peaks and the microprocessor monitors all leakage current pulses in each half cycle (8.33 milliseconds) for all three phases and provides a data summary every 9 hours. The data contain the maximum leakage current pulses, have the accumulated peaks classified into each of the five ranges, and have autonomy for 6 months, after which the data need to be downloaded and the system reset to begin a new monitoring period [10].

C. Results of leakage current measurement

Samir Ahmed Al-Gailani and Ali Ahmed Salem measured leakage current on high voltage glass insulators with uniform/non-uniform pollution and in wet conditions. They create artificial pollution by mixing kaolin at 10, 40, and 120 g, mixed with one liter of distilled water and an appropriate amount of NaCl which is common salt, according to the *ESDD* (mg/cm^2) severity required. The contamination severity is categorized into four levels of pollution based on *ESDD* value as: low (0.03 mg/cm^2), moderate (0.09 mg/cm^2), high (0.15 mg/cm^2), and severe (0.24 mg/cm^2).

They experimentally measured leakage current and do its analysis. The characteristics of leakage current as peak value, harmonic components, harmonics index of the leakage current signal for the insulators were evaluated for identification and prediction of flashover or risk usage according to the conduction of the leakage current. The shape of the leakage current waveform was used to monitor insulators. There is a nonlinear impedance of the pollution layer on the insulator surface which can be used to evaluate leakage current index R_{HI} which identify changes of 3rd, 5th, and 7th harmonic components:

$$R_{HI} = \frac{I_3}{(I_5+I_7)} \quad (1)$$

Where I_3, I_5, and I_7 represent the amplitude of 3rd, 5th, and 7th harmonic components of leakage current.

The proposed index provided, can be used as an effective tool in predicting the contamination severity on outdoor insulators. The increase of the leakage current index influences the decreased flashover voltage gradient [11].

Next researchers who use leakage current to study the performance of insulators in polluted and wet conditions are Dini Fauziah, Heldi Alfiadi, Rachmawati, Suwarno. By leakage current measurement they prove that silicone rubber coating can improve the performance of ceramic insulators due to better hydrophobicity. 6 silicone rubber coated, and 6 uncoated 20 kV distribution insulators have been installed in the field to get a natural aging process for 5 years and 4 months. The experiment was located in a coastal area that is heavily polluted by sea salt. Measurement of leakage current which flows on the insulator surface was done by measuring the voltage across a series resistance using a Digital Oscilloscope. Leakage current waveforms including low and high-frequency components were obtained. So fast Fourier transform was conducted to find out harmonic content and calculate Total Harmonic Distortion (*THD*):

$$THD = \frac{\sqrt{\sum_{n=2}^{\infty} I_n^2}}{I_1} \quad (2)$$

Where I_n represents harmonic components and I_1 is fundamental current.

The results show that the leakage current of both coated and uncoated insulators increase and decrease with the change of environmental condition. Humidity degree affected leakage current magnitude while THD was affected by the distribution of pollutants at the insulator's surface. Morning time was a susceptible condition to flashover occur because it has the average highest humidity, so it was important to monitor the condition of the insulator to avoid unexpected flashover. Silicone rubber coating can increase hydrophobicity and surface thermal capacity of ceramic insulators so it can be a solution to improve its performance especially for heavily polluted areas with high humidity. After the experiment in the coastal area, they prove that silicone rubber coated ceramic insulator has better performance [12].

A. Salem, R. Abd-Rahman, Samir A. Al-Gailani create a formulation of a new severity of harmonic index refers to a ratio of the sum of 5th and 7th to the 3rd harmonics component of leakage current. Laboratory tests were conducted on glass and porcelain insulators in dry and wet conditions with artificial

pollution created by salt-fog. Harmonic index $K_{(5+7)/3}$ is defined as follows:

$$K_{(5+7)/3} = \frac{(5^{th}+7^{th})}{3^{rd}} \quad (3)$$

Where 5th, 7th, and 3rd represent the fifth, seventh, and third leakage current harmonics components.
After measurement in dry and wet conditions with pollution levels represent light (*ESDD* = 0,06 mg/cm^2), medium (0,12 mg/cm^2), and heavy (0,2 mg/cm^2) pollution they successfully proposed and verified a new technique based on harmonics component of leakage current which is referred to as $K_{(5+7)/3}$ indicator, where they correctly recorded 90 from 96 tests. The proposed index, $K_{(5+7)/3}$ can be effectively used as a tool to predict the pollution severity of an outdoor insulator. The insulator is considered under a normal and clean condition if $K_{(5+7)/3} > 3$ %. Next, pollution severity is classified as moderate if 2 % < $K_{(5+7)/3} < 3$ %, while the pollution level is considered to be in heavy or extreme condition if $K_{(5+7)/3} < 2$ %. The insulator is predicted to be in a critical condition when $K_{(5+7)/3} < 1.7$ % whereby the probability of flashover occurrence is high [13].

R. Gulasekara and P. Subburaj measured flashover voltage and leakage current at a rated frequency of 50 Hz and the ambient temperature of 30 ˚C on a porcelain cap-and-pin insulator where contamination layer was made by bird excretion. They measured in wet and dry conditions.
Results show that wet pollution from birds decreases flashover voltage by 28 %. Leakage current increase in dry and polluted conditions by 14 %. This research presents the basic changes of insulator performance when the insulator is wet and polluted and how any kind of pollution can affect the insulator´s electric strength [14].

V. NUMERICAL SIMULATION MODELS

Simulated studies had become important as studies with an experimental component. Recent studies with numerical simulation models of insulators show us that numerical techniques are valid to calculate the electric field and potential distribution on insulators such as the Boundary Element Method (BEM), the Finite Difference Method (FDM), and the Finite Element Method (FEM) [15].

A. Finite Element Method (FEM)

Generally, in engineering, the FEM has already been established for a long time and is a successful numerical method to solve electric field problems, using the discretization of the domain, as well as coupling with other physical processes.
C. Mateus, F. A. Barata and R. Luís studied the Effects of Broken Skirts and Pollution on Voltage Distribution for Cap and Pin Glass Insulators (I160BS), where they create a mathematical model in finite element software (FEMM©).
After simulation of voltage distribution profile across fourteen units of glass insulators string with stubs in clean and in polluted conditions, results show that the pollution presence alters and exacerbates the stubs effect in the voltage distribution profile along the glass insulator string, being the worst-case for total polluted conditions compared to clean conditions and partial pollution [15].

Fig. 3 Glass insulator U160BS [15].

B. Boundary Element Method (BEM)

Boundary Element Method (BEM) is a numerical method to solve the partial differential equations which are represented in boundary integral equations. It is based on the principle of weighted residuals, where the fundamental solution is chosen as the weighing function. According to the method, the boundary is discretized into many elements over which the weighing function is evaluated by error minimizing of weighted residuals [16].
M. Ramesha, L. Cui, R.S. Gorur presents measurements and calculations of an electric field in samples with deliberately planted superficial and internal defects. Composite insulators and cylindrical rods of the housing material was used. The goal of this work is to find out indicators of impending failure by measuring and calculating the viability of electric field of composite insulators that can develop such defects during manufacture and service. Measurements were performed with an Electro-optic probe based on the Pockels effect. Calculations were performed using a 3-dimensional Boundary Element Method package.
After the electric field on composite insulators and cylindrical samples with different kinds of defects was measured and computed, results show good agreement between measurements and computational method. Tests represent various defects that can occur in composite line insulators. Defects change the electric field beside healthy insulator and decrease as the distance between the probe and defect increases. Also, they refer that Defects that are internal or nonconducting do not produce a significant change in the external electric field [16].

C. Finite Difference Method (FDM)

Finite Difference Method (FDM) provides the approximate solution of the partial differential equation. The aim of the FDM is to replace the electromagnetic field problems in the continuous domain with a discretized domain with a finite number of regular nodes [17].
As in [12] M. Ramesh and R.S. Gorur study composite insulators with and without defects. An electric field is calculated by a 2-dimensional stretched grid finite-difference to examine composite insulators up to 1000 kV. This method is shown to be applicable for composite insulators with defects. Different cases, such as single string, double string with and without corona, and grading rings were evaluated for healthy insulators. A conducting defect was simulated at the high voltage electrode to evaluate 69 and 138 kV insulators. The electric field computed for 69 kV composite insulator with a conducting defect at the high voltage electrode was compared with the actual measurements [17].

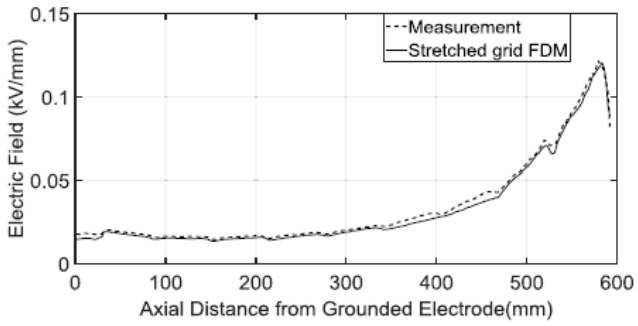


Fig. 4 Comparison of the measured and calculated electric field, [17].

Figure 4 shows us that experimental methods have comparable results as numerical simulation using the finite difference method [17].

VI. The Summary of Solved and Unsolved Problems

Nowadays the biggest problem of insulator´s performance is how they withstand industrial and environmental stresses in form of pollution and rainy, foggy, dusty, snowy weather. That's why researchers continuously measure stress influence on insulators.

Basic researches monitored insulators by application of AC and DC breakdown voltage as well as impulse voltages that can be generated by switching processes. Results of this measurement can give us fundamental values about insulators withstand voltage level in different environmental conditions and to compare which type of insulator is suitable for which environment. Results show us that composite insulators are beneficial if we compare them with porcelain insulators of the same type because of their electric strength which was higher in cases where AC and switching impulse voltages were used and reach the same electric strength in case of atmospheric impulse voltage. Question is, how their electric strength changes after a long period when aging occurs [6].

The further the more numerical simulation models are popular because of the software improvement that can simulate and calculate almost every situation in almost every condition. A wide range of modeling programs is available to create a very accurate copy of a real insulator. Then it is possible to create punctures, insulator surface damage, pollution layer with the accurate specification, and simulation of humidity and wet layer. The most frequently used is the finite element method which has already been established for a long time and is a successful numerical method to solve electric field problems. The advantages of FEM are the capability of handling complex geometry. The second is the finite difference method (FDM) which provides the approximate solution of the partial differential equation. If we compare FEM and FDM, FDM is simpler than FEM if we need to solve fewer complex problems. Boundary Element Method (BEM) is a numerical method to solve the partial differential equations which are represented in boundary integral equations. That means this method is advantageous when a small surface is needed to be studied. So appropriate numerical method for insulators depends on the complexity and conditions in which is insulator investigated [15], [16], [17].

The last measuring method that is described in this article is the measurement of leakage current. Measurement of leakage current is used to indicate a danger of failure of an insulator. Leakage current that is measured on the surface of the insulator can help us to estimate contamination level taking into account the humidity of the polluted layer, of the insulator that decreases the value of resistance. Researchers start to analyze the harmonic content of leakage current to find the way how to predict contamination level and evaluate the decrease of flashover voltage gradient. Two leakage current indexes are mentioned in the article, leakage current index R_{HI} and Harmonic index $K_{(5+7)/3}$. Both can predict the contamination severity on outdoor insulators and help us to prevent failures.

My research will continue with the next way how to use leakage current as an indicator of contamination level. We will measure leakage current in the wide range of voltage frequency with a constant value of voltage and check how frequency changes will influence the value of leakage current in different humidity and pollution level.

References

[1] M. Farzaneh and W. A. Chisholm, *Insulators for icing and polluted environments*. Piscataway, NJ : Hoboken, NJ: IEEE Press ; J. Wiley, 2009.

[2] B. Cao, L. Wang, and F. Yin, "Measurement of Saturated Water Absorption of the Contamination Layer Deposited on Insulator Surface," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10804–10811, Nov. 2019, doi: 10.1109/JSEN.2019.2932192.

[3] S. Hardi, Y. Tarigan, H. Zulkarnaen, and A. Hasibuan, "Influence of Artificial Pollutants on Disc Insulators under Dry and Wet Conditions on Leakage Current and Flashover Voltage," in *2019 3rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, Medan, Indonesia, Sep. 2019, pp. 174–178, doi: 10.1109/ELTICOM47379.2019.8943858.

[4] M. Ramesh and R. S. Gorur, "Stretched grid finite difference method for computation of electric field in composite insulators with defects," *Electric Power Systems Research*, vol. 192, p. 106875, Mar. 2021, doi: 10.1016/j.epsr.2020.106875.

[5] S. Venkataraman and R. S. Gorur, "Prediction of flashover voltage of non-ceramic insulators under contaminated conditions," *IEEE Trans. Dielect. Electr. Insul.*, vol. 13, no. 4, pp. 862–869, Aug. 2006, doi: 10.1109/TDEI.2006.1667747.

[6] K. A. Aravind, P. Rajamani, B. Krishna, K. Sandhya, and P. M. Nirgude, "Flashover Performance of Composite and Porcelain Insulators," in *2019 International Conference on High Voltage Engineering and Technology (ICHVET)*, Hyderabad, India, Feb. 2019, pp. 1–3, doi: 10.1109/ICHVET.2019.8724328.

[7] L. Zhengfa *et al.*, "Study on leakage current characteristics and influence factors of 110kV polluted composite insulators," in *2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM)*, Xi'an, May 2018, pp. 896–900, doi: 10.1109/ICPADM.2018.8401173.

[8] S. Deb, R. Ghosh, S. Dutta, S. Dalai, and B. Chatterjee, "Effect of humidity on leakage current of a contaminated 11 kV Porcelain Pin Insulator," in *2017 6th International Conference on Computer Applications In Electrical Engineering-Recent Advances (CERA)*, Roorkee, Oct. 2017, pp. 215–219, doi: 10.1109/CERA.2017.8343329.

[9] J. Dadashizadeh Samakosh and M. Mirzaie, "Analysis of leakage current characteristics during aging process of SiR insulator under uniform and longitudinal non-uniform pollution conditions," *Measurement*, vol. 147, p. 106862, Dec. 2019, doi: 10.1016/j.measurement.2019.106862.

[10] I. Ramirez, R. Hernandez, and G. Montoya, "Measurement of leakage current for monitoring the performance of outdoor insulators in polluted environments," *IEEE Electr. Insul. Mag.*, vol. 28, no. 4, pp. 29–34, 2012, doi: 10.1109/MEI.2012.6232007.

[11] A. A. Salem *et al.*, "Risk Assessment of Polluted Glass Insulator Using Leakage Current Index Under Different Operating Conditions," *IEEE Access*, vol. 8, pp. 175827–175839, 2020, doi: 10.1109/ACCESS.2020.3026136.

[12] R. G. Pandian and P. Subburaj, "Leakage current and flash over performance analysis of 11 kv pin insulator under Bird excretion pollution," in *2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS)*, Nagercoil, India, Apr. 2016, pp. 311–314, doi: 10.1109/ICEETS.2016.7583771.

[13] A. A. Salem, R. Abd-Rahman, S. A. Al-Gailani, M. S. Kamarudin, H. Ahmad, and Z. Salam, "The Leakage Current Components a Diagnostic Tool to Estimate Contamination Level on High Voltage

Insulators," *IEEE Access*, pp. 1–1, 2020, doi: 10.1109/ACCESS.2020.2993630.

[14] D. Fauziah, H. Alfiadi, Rachmawati, and Suwamo, "The effect of coating on leakage current characteristic of coast field aged ceramic insulator," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Yogyakarta, Sep. 2017, pp. 1–7, doi: 10.1109/EECSI.2017.8239140.

[15] C. Mateus, F. A. Barata, and R. Luis, "Effects of Broken Skirts and Pollution on Voltage Distribution for Cap and Pin Glass Insulators," in *2020 IEEE 14th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, Setubal, Portugal, Jul. 2020, pp. 30–35, doi: 10.1109/CPE-POWERENG48600.2020.9161644.

[16] M. Ramesh, L. Cui, and R. S. Gorur, "Impact of superficial and internal defects on electric field of composite insulators," *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 327–334, Mar. 2019, doi: 10.1016/j.ijepes.2018.10.013.

[17] Ali. A. Salem, R. A. Rahman, M. S. Kamarudin, and N. A. Othman, "Factors and models of pollution flashover on high voltage outdoor insulators: Review," in *2017 IEEE Conference on Energy Conversion (CENCON)*, Kuala Lumpur, Malaysia, Oct. 2017, pp. 241–246, doi: 10.1109/CENCON.2017.8262491

A survey on deep learning based pathological speech detection

[1]*Máté HIREŠ (2nd year),*
Supervisor: [2]*Peter DROTÁR*

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

[1]mate.hires@tuke.sk, [2]peter.drotar@tuke.sk

***Abstract*—This paper presents an overview of the most frequently used datasets that contain pathological speech samples. A survey of related best resulting works on deep learning approaches for pathological speech detection or classification is also provided. A brief analysis of the mentioned databases is included as well.**

***Keywords*—Deep learning, Parkinson's disease, Pathological speech, Speech processing**

I. Introduction

Pathological voice can be caused by many neurodegenerative or respiratory system disorders. The most common diseases affecting the human voice are the Parkinson's disease, lung cancer, cystic fibrosis, nodules, polyps and others. Since the detection of such diseases is very important to prevent other chronical lesions, automatic diagnosis systems has received more attention. The traditional diagnostic methods are more expensive and time-consuming. In the recent years, the machine learning based speech processing methods became more commonly used. Since then, many accurate methods have been developed with notable results achieved. The most important part of the pathological speech detection process is the appropriate choice of features. Mekyska, et al. [1] provided a set of 92 speech features with description. They verified the importance of the given features on three different datasets: the Massachusetts Eye and Ear Infirmary Voice Disorders Database (MEEI) [2], Príncipe de Asturias Database (PdA) [3] and the Czech Parkinsonian Speech Database (PARCZ) [4]. Pishgar, et al. [5] proposed a classification model based on the Mel-frequency Cepstral coefficients (MFCC) and the MFCC delta features extracted from the speech signals. They tested their model on the Far Eastern Memorial Hospital database (FEMH) [6]. Orozco-Arroyave, et al. [7] analyzed the spectral-cepstral features of the data to classify the speech pathologies for the patients from the FEMH and the PC-GITA database [8]. Martinez, et al. [9] pointed out the importance of noise-related features while classifying the pathologies in the MEEI and the Saarbruecken Voice Database (SVD) [10].

The aim of this paper is to provide an overview on some the existing pathological speech datasets and works dealing with pathological speech processing using deep learning methods.

II. Pathological voice datasets

This section provides an overview of the most frequently used datasets containing pathological speech samples.

A. Saarbruecken Voice Database

The SVD dataset is the most comprehensive dataset containing pathological speech. It consists of 2042 samples, from which 687 are from healthy and 1355 from pathological subjects. The samples are recorded at 50 kHz rate and 16 bit resolution. Every session consists of three tasks:

1) sustained phonation of the vowels [i, a, u] with normal, high and low intonation,
2) sustained phonation of the vowels [i, a, u] with rising-falling intonation,
3) Reading of the sentence "Guten Morgen, wie geht es Ihnen?".

This databse includes 71 different pathologies [10].

B. The Massachusetts Eye and Ear Infirmary Voice Disorders Database

The MEEI database is the most frequently used dataset containing pathological speech samples. It consists of two sessions:

1) sustained phonation of the vowel /a/
2) reading the text of the "The Rainbow Passage"

The first session includes 657 pathological and 53 healthy samples, the second session contains 661 pathological and 53 healthy samples. The healthy samples are recorded at 50 kHz, while the pathological samples are recorded at either 10, 25 or 50 kHz [2], [1].

C. Far Eastern Memorial Hospital database

The FEMH database consist of 60 healthy and 402 pathological voice samples, which include voice pathologies, such as vocal nodules, polyp, cyst, glottic neoplasm, vocal atrophy, laryngeal dystonia, unilateral vocal palsy and sulcus vocalis. The dataset contains recordings of sustained phonation of the vowel /a/. The recordings are sampled at 44,1 Hz and 16 bit resolution [6].

D. PC-GITA

This database includes 100 samples: 50 samples from patients with Parkinson's disease (PD) and 50 from healthy subjects. The samples were recorded at 44100 Hz sampling rate and 16 bit resolution. One recording includes five different tasks: (1) sustained phonation of vowels /a/, /e/ /i/, /o/, /u/; (2) Reading of 25 different Spanish words; (3) Reading of a dialog; (4) Reading of sentences with different emphasis in some marked words; (5) A spontaneous speech [8].

E. Príncipe de Asturias Database

The PdA database was recorded at the Hospital in Alcalá de Henares, Madrid. It consists of 439 speech recordings, from which 239 are healthy and 200 pathological. The present pathologies are polyps, nodules, carcinomas and oedemas. The PdA dataset contains one recording per subject, which is the sustained phonation of the vowel /a/ [3].

F. Czech Parkinsonian Speech Database

The PARCZ database contains voice samples from 52 healthy and 57 pathological speakers, which were diagnosed with Parkinson's disease. The recordings contain 91 different tasks to analyse the speech disorders caused by PD. The samples are recorded at 48 kHz. The detailed description of the speech tasks are available in [4].

III. Results

The following table contains a list of related articles with the best achieved results in the field of pathological voice detection and classification. All the proposed approaches use deep neural networks to achieve the results. On the SVD we were able to reach 74 % testing accuracy using only the /a/ vowel subset. We proposed another deep learning approach, reaching 94,33 % accuracy using the /u/ vowel subset of the PC-GITA dataset [11].

Reference	Database	Best accuracy
Fang, S., et al. [6]	MEEI	99,33 %
Lee, J., et al. [12]	MEEI	99,3 %
Hireš, M. [11]	PC-GITA	94,33 %
Fang, S., et al. [6]	FEMH	94,26 %
Chuang, Z. Y., et al. [13]	FEMH	93,1 %
Vasquez-Correa, J. C., et al. [14]	Multimodal data	92,3 %
Wodzinski, M., et al. [15]	PC-GITA	91,7 %
Vasquez-Correa, J. C., et al. [16]	PC-GITA	85,6 %
Vasquez-Correa, J. C., et al. [16]	PARCZ	89,2 %
Lee, J., et al. [12]	SVD	75,18 %
Hireš, M. [11]	SVD	74 %
Harar, P., et al. [17]	SVD	71,36 %
Wu, H., et al. [18]	SVD	71 %

TABLE I
List of some articles related to deep learning based pathological speech detection or classification

IV. Conclusion

Even, when the best results were obtained on the MEEI and FEMH datasets, they contain imbalanced data. Furthermore, the recordings of the MEEI dataset are sampled in three different environments. The results obtained on these datasets may not be accurate.

The SVD dataset contains the most data for pathological speech detection and classification. However, its data is also imbalanced. The PARCZ and the PC-GITA datasets contain only a few samples (<110), which is insufficient to train an accurate model on. Moreover, the PARCZ dataset is focusing rather on older people, the average age of the speakers is approximately 65. Another hindering factor is the availability of the data. From the datasets mentioned in this work, only the SVD and the PC-GITA dataset is freely available.

A sufficient amount of well balanced freely available speech data could become a milestone in the field of pathological speech processing. Research work in this field has a wide area based on its applicability. Working on specific speech dysfunction, analysing other acoustic features, applying new methods of speech processing or development of standardised databases could provide improvements in this field.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under contract No. APVV-16-0211 and 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 contract VEGA 1/0327/20.

References

[1] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy *et al.*, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94–111, 2015.

[2] Massachusetts Eye and Ear Infirmary, "Voice disorders database, version. 1.03 (cd-rom)," *Lincoln Park, NJ: Kay Elemetrics Corporation*, 1994.

[3] J. I. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, I. Cobeta-Marco, R. González-Herranz, and C. Ramírez-Calvo, "Acoustic analysis of voice using wpcvox: a comparative study with multi dimensional voice program," *European archives of oto-rhino-laryngology*, vol. 265, no. 4, pp. 465–476, 2008.

[4] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilová, and I. Rektorova, "Motor aspects of speech imparment in parkinson's disease and their assessment," *CESKA A SLOVENSKA NEUROLOGIE A NEUROCHIRURGIE*, vol. 74, pp. 662–668, 01 2011.

[5] M. Pishgar, F. Karim, S. Majumdar, and H. Darabi, "Pathological voice classification using mel-cepstrum vectors and support vector machine," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec 2018, pp. 5267–5271.

[6] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *Journal of Voice*, vol. 33, no. 5, pp. 634–641, 2019.

[7] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolanos, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hönig, and E. Nöth, "Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1820–1828, 2015.

[8] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rativa, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease." in *LREC*, 2014, pp. 342–347.

[9] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the saarbrücken voice database with calibration and fusion of scores using multifocal toolkit," in *Advances in Speech and Language Technologies for Iberian Languages*. Springer, 2012, pp. 99–109.

[10] W. J. B. M. Pützer, "Saarbruecken voice database." [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de

[11] M. Hireš, "Advanced speech processing methods for pathological speech detection," Ph.D. dissertation, Technical University of Košice, 2021.

[12] J. Lee and H.-J. Choi, "Deep learning approaches for pathological voice detection using heterogeneous parameters," *IEICE Transactions on Information and Systems*, vol. 103, no. 8, pp. 1920–1923, 2020.

[13] Z.-Y. Chuang, X.-T. Yu, J.-Y. Chen, Y.-T. Hsu, Z.-Z. Xu, C.-T. Wang, F.-C. Lin, and S.-H. Fang, "Dnn-based approach to detect and classify pathological voice," in *2018 IEEE international conference on big data (big data)*. IEEE, 2018, pp. 5238–5241.

[14] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of parkinson's disease: a deep learning approach," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1618–1630, 2018.

[15] M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave, and E. Nöth, "Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 717–720.

[16] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Convolutional neural network to model articulation impairments in patients with parkinson's disease." in *INTERSPEECH*, 2017, pp. 314–318.

[17] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget, and Z. Smekal, "Voice pathology detection using deep learning: a preliminary study," in *2017 international conference and workshop on bioinspired intelligence (IWOBI)*. IEEE, 2017, pp. 1–4.

[18] H. Wu, J. Soraghan, A. Lowit, and G. Di Caterina, "A deep learning method for pathological voice detection using convolutional deep belief networks," *Interspeech 2018*, 2018.

Feature selection based on neural networks

[1]*Peter BUGATA (3rd year),*
Supervisor: [2]*Peter DROTÁR*

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

[1]peter.bugata@student.tuke.sk, [2]peter.drotar@tuke.sk

***Abstract*—Feature selection provides an effective way to solve the problem of high-dimensional data analysis by removing irrelevant and redundant data. It can improve learning performance and reduce computation time and demands on computational resources. This paper discusses the relationship between feature selection and neural networks and presents the possibilities of using neural networks for feature selection.**

***Keywords*—feature selection, high-dimensional data, neural network, neural-network-based feature selection**

I. Introduction

In recent years, a giant amount of data has been generated through many computer applications. Data obtained from social media and the growing Internet of Things and cloud computing often have an extremely large number of attributes (variables, features). These high-dimensional datasets pose a serious challenge for researchers in the field of data mining because when applying machine learning algorithms with tens or hundreds thousands of attributes, a phenomenon known as the curse of dimensionality occurs. This causes a noticeable reduction of the prediction performance of the algorithms originally designed for low-dimensional data. Moreover, high dimensionality increases demands on computational resources and prolongs computational time.

There are two means of reducing dimensionality: feature extraction (FE) and feature selection (FS). FE reduces dimensionality through mapping a high-dimensional feature space to a new low-dimensional space. Its drawback is that the newly constructed features have no physical meaning for interpretation. FS methods utilize certain criteria to evaluate the quality of features, based on which they select a smaller subset of relevant features from the original feature set, while other features evaluated as irrelevant or redundant are removed. FS can increase learning performance and save computational resources. It also contributes to a better understanding of the learning model and data.

In recent years, much attention has been paid to artificial neural networks and deep learning, which have been successful in solving many current machine learning problems, such as image and speech recognition and processing, machine language translations, games, and medical diagnostics. An interesting question is the impact of FS on the performance of neural networks, but also the use of neural networks and deep learning for FS.

This paper focuses mainly on FS methods based on neural networks. The rest of the paper is organized as follows: Section II discusses the relationship between FS and neural networks. Section III presents some FS approaches based on neural networks and examples of their use. Section IV summarizes our conclusions and outlines further research.

II. Neural Networks and Feature Selection

Deep learning is one of the major advances in machine learning in recent years. Deep learning algorithms construct new abstract high-level features from the original low-level variables and form distributed representations of data [1]. Consider an arbitrary feedforward neural network with m input neurons, i.e., original input variables. If h is the number of neurons of any hidden layer, then the output of this layer presents a complex transformation from the original m-dimensional space to the h-dimensional space. When $m \gg h$, it is a dimensionality reduction and the neural network can be used to transform the original dataset to a new dataset in a lower dimensional space.

The built-in FE ability of neural networks may be even more interesting if the input variables have a certain structure. For example, a photograph or recording of an audio signal can first be processed by a convolutional neural network and converted to a multidimensional vector of real numbers representing the values of new abstract features. Time series can sometimes be interpreted as sequences of different lengths. Such sequences can be converted to vectors of real numbers of a fixed dimension using a recurrent neural network.

Although it seems that advanced machine learning can be done without the use of FS methods, in some cases, when the number of data points is not large enough, deep learning should be combined with FS to obtain better learning performance [2]. Moreover, irrelevant features might cost a great deal of resources during the neural network learning process, and removing them can save training time [1]. FS reduces the complexity of the model at the input level and makes its interpretation easier.

On the other hand, because the weights of irrelevant features will be close to zero during neural network training, deep learning can also be used to FS. Deep-learning-based feature selection as well as feature-selection-based deep learning are approaches that should be further investigated [1].

III. Feature Selection Based on Neural Networks

Let $\mathcal{F} = \{X_1, X_2, \ldots, X_m\}$ be the set of m explanatory variables or features, $\{x_1, x_2, \ldots, x_n\}$ be the set of n observations of a dataset $\boldsymbol{X} \in \mathbb{R}^{n \times m}$, and let $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)$ be the target variable. The goal of FS is to find a set $\mathcal{S} \subset \mathcal{F}$ of p features, $p \ll m$, that optimally characterizes the target variable $\boldsymbol{Y}$. Based on the features in $\mathcal{S}$, a new dataset $\boldsymbol{X}^* \in \mathbb{R}^{n \times p}$ is extracted, which keeps most of the information about $\boldsymbol{X}$. The quality of the selection is assessed according to various criteria. The most important of them is usually considered to be the impact of the selection on the performance of classifiers.

A. Determining Salient Input Variables

Some FS methods are based on the assumption that the weights of the trained neural network corresponding to irrelevant features are close to zero. To determine the importance of features for making a correct prediction, they define various saliency metrics that involve the weights of the network or the derivative of the network error function, or both [3]. A simple example is the sum of the squared weights between a given input variable and the first hidden layer. The values of these saliency metrics are calculated from the network trained with all input variables, and based on them, the variables are ranked from most significant to least significant.

Several FS algorithms sequentially eliminate irrelevant variables according to some criterion evaluated in each step for the remaining variables. In [4], the importance of a variable is measured by the difference of the values of the objective function before and after its removal. In [5], variables are excluded based on the prediction accuracy after their removal.

B. Sparse FS Using Neural Networks

Another FS methods try to reduce the number of non-zero weights during network training using appropriate regularization techniques, and thus obtain a smaller number of the most important features. The idea of regularization is to add a penalty term to the objective function minimized during the process of learning network parameters.

Let us consider a feedforward neural network with H hidden layers in the model. We denote the set of model parameters by $\boldsymbol{\theta} = \{\boldsymbol{W}^1, \boldsymbol{b}^1, \ldots, \boldsymbol{W}^{H+1}, \boldsymbol{b}^{H+1}\}$, where $\boldsymbol{W}^h = (w_{ij}^h)$ is the matrix of weights of the connections between the $(h-1)$-th and h-th layers and $\boldsymbol{b}^h$ is the bias in the h-th layer for $h = 1, \ldots, H+1$. The regularized objective function has the form:

$$f(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \lambda \sum_{h=1}^{H+1} \Omega(\boldsymbol{W}^h), \tag{1}$$

which is the sum of the loss function $L(\boldsymbol{\theta})$ and the regularization term $\sum_{h=1}^{H+1} \Omega(\boldsymbol{W}^h)$ multiplied by the regularization parameter $\lambda \in R_0^+$, which controls the balance between the loss term and the regularization term.

Many well-known regularizers are based on the p-norm, defined for the vector $\boldsymbol{w} = (w_1, \ldots, w_n)$ and $p \in R_0^+$ as follows:

$$\|\boldsymbol{w}\|_p = \left(\sum_{i=1}^{n} |w_i|^p \right)^{\frac{1}{p}}. \tag{2}$$

Sparse regularization causes the reduction of many weights to zero during network training. The most used sparse regularizer is ℓ_1 based on the $\|\boldsymbol{w}\|_1$ norm. This is a convex approximation of the ℓ_0 regularizer based on $\|\boldsymbol{w}\|_0$ norm, which determines the number of non-zero weights. Although ℓ_0 appears to be ideal, it has been shown that solving an ℓ_0-regularized learning problem is NP-hard. To improve ℓ_1, some other non-convex metrics have been introduced, for example ℓ_p based on the $\|\boldsymbol{w}\|_p$ norm for $p \in (0, 1)$. Another alternatives are combinations of ℓ_1 and ℓ_2 (based on $\|\boldsymbol{w}\|_2$ norm), such as ℓ_{1-2}, which is the difference of ℓ_1 and ℓ_2, and their convex combination known as the elastic-net regularization defined for the vector $\boldsymbol{w}$ and $\alpha \in [0, 1)$ by

$$\alpha \|\boldsymbol{w}\|_2^2 + (1-\alpha)\|\boldsymbol{w}\|_1. \tag{3}$$

Neural-network-based FS methods presented in [6], [2] select important input variables of a feedforward neural network utilizing the ℓ_1 and $\ell_{1/2}$ regularization. In the FS method introduced in [7], a sparse one-to-one FS layer is included between the input layer and the first hidden layer. The network is trained using the elastic-net regularization.

C. Applications of Neural-Network-Based FS Methods

Several works introduce neural-network-based FS methods and their successful applications in bioinformatics [7], [8]. Deep-learning-based FS approaches are presented in [9], [10], where they are used for selection of important features in the context of action recognition from video records and remote sensing scene recognition.

IV. Conclusion

FS is an important pre-processing step of machine learning algorithms. The success of neural networks in solving many nonlinear learning problems has led to the question of their relationship with FS. This paper shows that FS can be useful for neural networks and also neural networks can be the basis for FS methods. The paper presents some neural-network-based FS approaches.

Further research will focus on the development of the new neural-network-based FS method proposed in [11], which selects important features using a special sparse FS layer.

Acknowledgment

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 contract VEGA 1/0327/20 and by the Slovak Research and Development Agency under contract No. APVV-16-0211.

References

[1] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70 – 79, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218302911

[2] J. Li and H. Liu, "Challenges of feature selection for big data analytics," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 9–15, 03 2017. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=arnumber=7887649

[3] L. M. Belue and K. W. Bauer, "Determining input features for multilayer perceptrons," *Neurocomputing*, vol. 7, no. 2, pp. 111–121, 03 1995.

[4] E. Romero and J. M. Sopena, "Performing feature selection with multilayer perceptrons," *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 431–441, 03 2008.

[5] Rudy Setiono and Huan Liu, "Neural-network feature selector," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 654–662, 05 1997.

[6] K. Sun, S. Huang, D. S. Wong, and S. Jang, "Design and application of a variable selection method for multilayer perceptron neural network with lasso," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 6, pp. 1386–1396, 2017.

[7] Y. Li, C.-Y. Chen, and W. W. Wasserman, "Deep feature selection: Theory and application to identify enhancers and promoters," in *Research in Computational Molecular Biology*, T. M. Przytycka, Ed. Cham: Springer International Publishing, 2015, pp. 205–217.

[8] R. Ibrahim, N. A. Yousri, M. A. Ismail, and N. M. El-Makky, "Multi-level gene/MiRNA feature selection using deep belief nets and active learning," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 3957–3960.

[9] D. Roy, K. S. R. Murty, and C. K. Mohan, "Feature selection using deep neural networks," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 07 2015, pp. 1–6.

[10] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 11 2015.

[11] P. Bugata and P. Drotar, "Feature selection based on sparse neural network layer with normalizing constraints," 2020. [Online]. Available: https://arxiv.org/abs/2012.06365

System for data collection from a CNC milling machine for the purpose of making a thermal model

[1]*Branislav FECKO (3rd year)*
Supervisor: [2]Tibor VINCE

[1,2]Department of Theoretical and Industrial Electrical Engineering, FEI TU of Košice, Slovak Republic

[1]branislav.fecko@tuke.sk, [2]tibor.vince@tuke.sk

***Abstract* — The paper describes s system for data collection from a CNC milling machine. Data collection is performed for the purpose of making a thermal model. The first part contains the definition of the concept of the entire system. The method of communication between the individual nodes of the system is also explained. The next part deals with the description of the graphical interpretation of collected data and the explanation of the need to collect certain data.**

***Keywords* — actuator, CNC machine, data collection, Grafana, InfluxDB, Python, temperature, thermal model, visualization**

I. Introduction

Proper temperature monitoring is required to understand the individual events and determine the effect of temperature. There are currently a number of sensors and measuring instruments for measuring temperature. But if you want to carry out a detailed analysis of the events affecting the heating of the system and the subsequent effect of temperatures on other parts of the system, it is necessary to create a robust database with the appropriate data[1]. The main information collected by our system includes temperature, stepper motors velocity, executed machine code, machine calibration parameters and milling information. Such an extensive database will allow us to research the effects of processes and heating in the system. Therefore, in this article we will focus on tools suitable for data collection and visualization.

II. The System Concept

The first main part of system is focused for temperature measurement. The design of this part consists of the InfluxDB database, the Grafana visualization tool, sensors DS18b20[1] and the ESP8266 microcomputer. The communication of the system between these nodes of the measuring system is ensured by the Mosquitto broker in combination with the MQTT communication protocol.

The next part provides the collection of data about movements, calibration parameters of the CNC machine and the current milling process. The collection of this data from the control program of the CNC milling machine(Machinekit) is implemented using the Python interface module. This module ensures direct communication with Machinekit and Python software, which send data to database InfluxDB. A diagram of the described system for data collection and visualization is shown in Fig. 1.

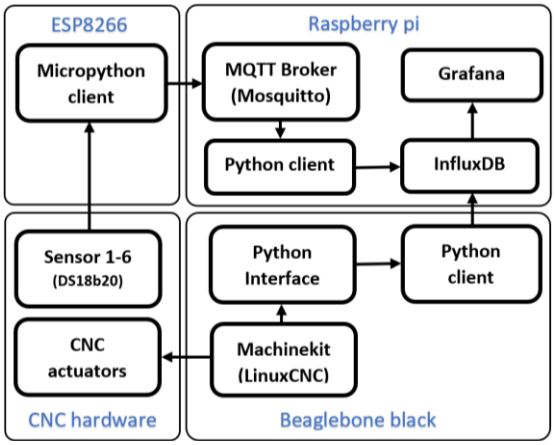


Fig. 1 The system diagram

Description of the Individual System Tools

Python and Micropython(Python version for ESP8266) programming languages. In Python are created programs for collecting data and sending them to the InfluxDB database.

MQTT is a standard messaging protocol for the Internet of Things (IoT). It is designed as an easy messaging for publishing and subscribing in which is ideal for connecting remote devices and is widely used in industry. The basis of MQTT communication is a publish / subscribe system. Each message is published with the topic name, which the broker sends to all clients who have subscribed to it[3].

InfluxDB is a time series database designed to handle high load during writing and querying. Contains dashboards, queries, tasks, and agents. In our case we use it to store all measured data. Subsequent data analysis requires a suitable selecting of data from the database, for this purpose was chosen Grafana visualization tool[4].

The last tool most important for visualization is the Grafana platform that allows to visualize data by creating graphs, dashboards and other visualization elements[5].

III. Parts of Graphical Visualization

Fig. 2 shows the layout of the Grafana visualization. The individual parts of it will be shown and described separately.

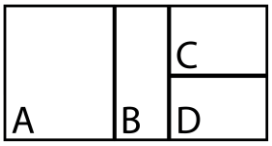


Fig. 2 Visualization layout in the Grafana environment

The first part of the visualization contains display of graphs. The graphs show temperature data and information about the current speed of the machine drives. Temperature visualization is needed due to future verification of the thermal model. The current speed of the drive is another necessary parameter for a more accurate analysis of iron losses, which are dependent on the speed of rotation of the motor. You can see the individual parameters depending on the time in Fig. 3.

Fig. 3 Graphs of temperatures and speeds of CNC machine drives (part A)

The input parameter for the future thermal model is the machine code of the CNC machine. This code consists of G-codes and M-codes, which are technical and geometrical information about the work process for the CNC machine. For the subsequent comparison of models and real machining, it is necessary to store this data in our database. You can see his interpretation in our Grafana visualization in Fig. 4.

Executed machine code

Time ↓	id	G_codes
2021-02-10 16:44:36	1032	G1Y4.516
2021-02-10 16:44:34	1031	G3X40.976Y-4.5...
2021-02-10 16:44:30	1030	G1X30.699Y-14....
2021-02-10 16:44:29	1029	G3X-30.687Y-14....
2021-02-10 16:44:28	1028	G1X-40.976Y-4.5...

Fig. 4 Tabular representation of executed machine code (part B)

The system stores the temperatures of all motors, the temperature inside the cover and the ambient temperature out of the cover. The current values of these temperatures are visualized in layer C. In its lower part there is information about the state of the CNC machine. Layer C of the Grafana visualization is shown in Fig. 5.

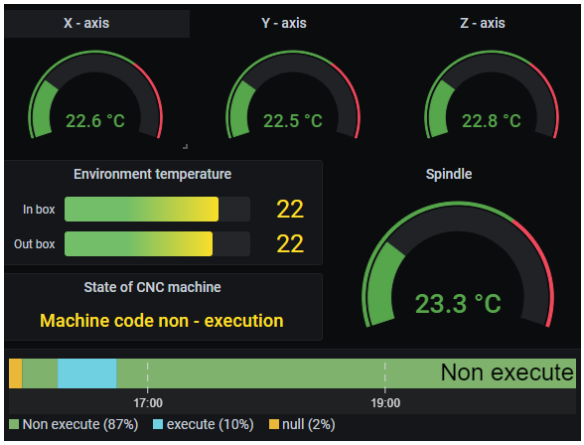


Fig. 5 Visualization of the current state of temperatures and machine operation (part C)

In the last part of the visualization, information about the machining process is displayed. Data such as workpiece material, tool used, maximum machine speed and drive calibration are factors influencing the heating temperature of the entire CNC machine. Therefore, for future analysis, it is necessary to be informed about these properties. Their visualization is shown in Fig. 6.

Milling information

tool	workpiece material	action info
none	none	no-load test

machine name	servo period	default velocity	max linear velocity
CNC Mill-M KTPF	1000000	20	25

Stepper motors calibration informations

axis	max_vel	max_acc	scale	dirsetup	dirhold	steplen	stepspace
AXIS_0	60	30	67	200	200	1000	1000
AXIS_1	60	30	67	200	200	1000	1000
AXIS_2	60	30	67	200	200	1000	1000

Fig. 6 Milling and machine calibration information (part D)

IV. Conclusion and Next Steps

The article described a comprehensive data collection system implemented by us for the purpose of manufacturing a thermal model of a CNC machine. The main advantage of the system is robustness and flexibility. The system also allows you to enter temperature prediction data and set alerts to parameters that should not exceed a certain value, such as temperature. The data stored, in the future can create a suitable set of data for a machine learning application and other methods of modeling.

References

[1] B. Fecko, T. Vince.: Design of temperature measurement system and data visualization. Journal of Industrial Electrical Engineering, Vol. 4, Issue 3 (2020) pp: 1-3, ISSN 2454-0900

[2] Dallas Semiconductor, DS18B20, [online] [cit: 10.2. 2021.] Available on the Internet: https://html.alldatasheet.com/html-pdf/58557/-DALLAS/DS18B20/181/1/DS18B20.html

[3] IoT Portal, MQTT, [Online] [cit: 10.2. 2021.] Available on the Internet: https://www.iot-portal.cz/2016/05/24/mqtt/

[4] InfluxData, InfluxDB, [Online] [cit: 12.2. 2021.] Available on the Internet: https://www.influxdata.com/products/ influxdb/

[5] Grafana Labs, Grafana, [Online] [cit: 12.2. 2021.] Available on the Internet: https://grafana.com/grafana

Impact of Photovoltaic systems and electric vehicles on a residential low voltage power grid

[1]*Dávid MARTINKO (4th year)*
Supervisor: [2]Michal Kolcun

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

[1]david.martinko@student.tuke.sk, [2]michal.kolcun@tuke.sk

***Abstract*— Study impacts of photovoltaic systems and electric vehicles in low voltage residential networks based on real historical smart meter data in suburban or rural area with household this time without photovoltaic systems and without use electric vehicles considering unbalanced loads across phases in 3 phase network feeder and 3 phase consumer connections.**

***Keywords*— Smart meter data, Low voltage, Load profile, Photovoltaic systems, Electric vehicle, Unbalanced network, Battery storage, Impact calculations**

I. Introduction

Installations of rooftop photovoltaic systems in households powered by low voltage local distribution network is gaining increasing popularity in recent years. Rooftop photovoltaic systems inject active power during the day at a time when there is usually a minimum level of household consumption. Therefore, in recent years, the authorities and system operators must have adopted several limitations on photovoltaics systems installations as maximum total system power or maximum load per phase.

Likewise, in recent years, the popularity of electric vehicles among the general population living in family house has increased due to some of their valuable options like reducing greenhouse gasses by offering low or free emissions transportation as an alternative to conventional vehicles. Charging and discharging activities of electric vehicles have a direct impact on the distribution network by step changes in power flow. One of the main impacts on distribution network power flow and overall local area load profile is a significant increase in power peaks because charging electric vehicles which usually takes place on the afternoon consumption peak after coming from work.

My work studies the integration and impacts of photovoltaic systems and electric vehicles in low voltage residential networks based on real historical smart meter data in a suburban or rural areas with households that time without photovoltaic systems or use of electric vehicles. from last year in rural in the form of single-phase slow charging during the night, excluding vehicle-to-grid capabilities.

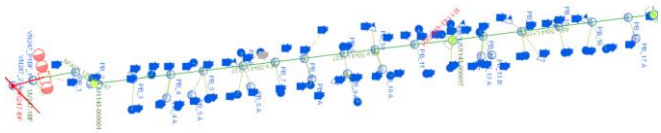


Fig. 1. Selected substation distribution feeder illustration

Several levels of vehicle ownership penetration and photovoltaics system installations with or without battery storage will be chosen to identify possible impacts to overall substation load profiles in the local power grid.

II. Methodology

The calculations are based on veritable smart meter data used by local distribution system operator (DSO) company Vychodoslovenska Distribucna a.s., This DSO operates on its own territory with circa 9500 MV/LV substations, own or foreign, with more than 660k electricity meters, of which is 95k smart meters. We choose a distribution substation with the highest smart meter ratio to a total number of electricity meters and the total number of connected households is more than 30, which represents a common suburban or rural area.

Distribution substation in suburban municipality Hriadky fed 55 customers with 54 installed smart meters. The 1 remaining household has insignificant annual consumption. The selected substation distribution feeder illustration is figured in Fig. 1. The selected test network is a 400 V 3 phase four-wire low voltage network with a radial feeder. It is to be noted that although this type of low voltage network is common in many parts of world except Northern America.

As the power grid is dynamic in nature, with its operation change from minutes-to-minutes, we subdivide load profiles time horizon into 15-minute time slots, t, such that

$t = \{1,2, \ldots, T\}$.

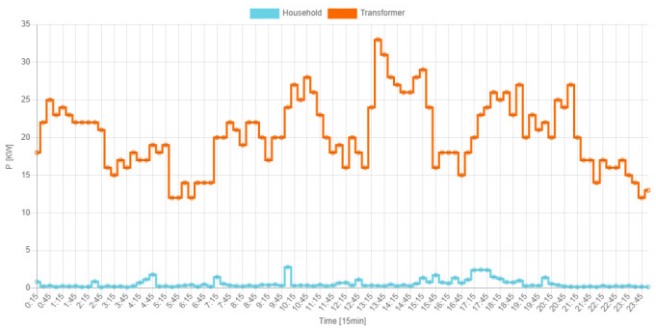


Fig. 2. Actual day load profile of transformer and randomly selected household

Total time step load profile of distribution station P_{sum} can be described as a sum of load profile P_i for each connected household i for time step t. Figured in Fig. 2.

$$P_{sum\,(t)} = \sum_{i=1}^{n} P_{i\,(t)}$$

According to the analysis of household load profiles, we can conclude, that those are simple loads without the usage of electric vehicles or photovoltaic systems, as we plan to simulate.

$$P_i = P_{PV} + P_b - P_{EV}$$

The available output power $P_{PV(t)}$ of photovoltaic system at maximum power point in time interval can be calculated based on global horizontal irradiance Q, photovoltaics panels area S_{PV}, system ambient temperature T, and photovoltaics system efficiency η [1].

$$P_{PV(t)} = \eta\, S_{PV} Q_{(t)}\,(1 - 0{,}005\,(T_{(t)} - 25))$$

In Fig. 3 is figured calculation of photovoltaic system power output profile based on historical data from Solcast [2] and technical parameters one of the most used solar panels [3] for one day at middle of summer and overall photovoltaic capacity is 3,36 kWp.

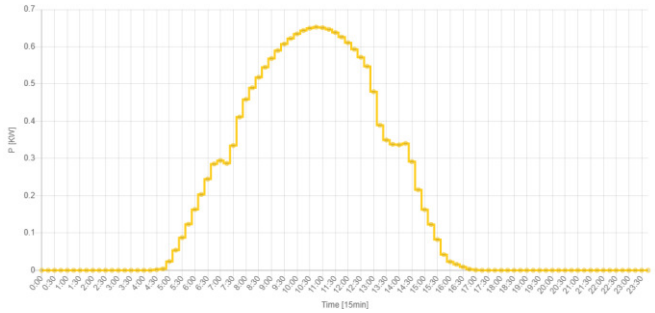


Fig. 3. Output profile of photovoltaic system in of middle summer

The sizing of simulated battery systems is based on available models for households with consideration of the economic efficiency of these solutions. Battery load or the state of charge of the battery can be calculated by the following equations [4] [5]

$$P_{b\,(t)\,discharge} = P_{b(t-1)}(1-\sigma) - \left(\frac{P_{PV(t)}}{\eta_i} - P_{(t)}\right)$$

$$P_{b\,(t)\,charge} = P_{b(t-1)}(1-\sigma) + \left(P_{PV(t)} - P_{(t)}/\eta_i\right)\eta_b$$

Where $P_{b(t)}$ and $P_{b(t-1)}$ represent battery load demand in adjacent time intervals, σ is the battery self-discharging factor and η_i, η_b are the efficiency of inverter and battery charging.

Power grid structure needs to be analyzed, developed, and improved due to the increased utilization of electric vehicles. Vehicle charging scenarios used in the simulation are real historical smart meter data from DSO fleet cars used by company workers. These charging systems are "slow", which means 1 phase about 4 kW and does not correspond to the actual mixed types of charging. Simulation of charging, defined by power load P_{EV}, will be based on various types of charging, 1 or 3 phases, slow or fast with the statistically expected use of vehicles during the days and the corresponding needs to recharge the battery [6].

Most analysis of the impact of photovoltaics system or electric vehicles load profiles is tied to balanced residential low voltage networks with 1 phase customer loads.

As we can see in Fig.4, in real conditions is low voltage network unbalance due to various loads across phases caused by 3 phase household connection of most suburban residential

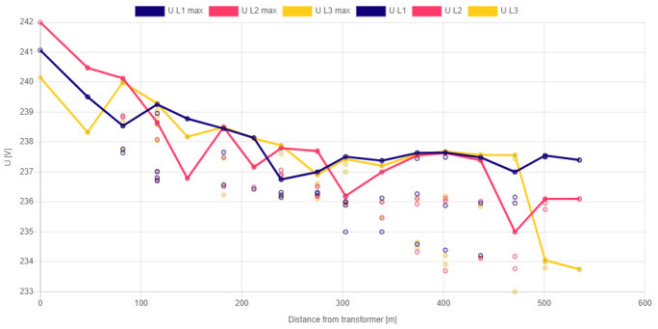


Fig. 4. Voltage unbalance changes along the network according to the distance from transformer in randomly selected time interval during day

houses. The individual points represent the measured effective voltage values in the individual smart meters arranged according to the distance from the transformer station. For a better illustration of the voltage drop in the network due to unbalanced load, the maximum measured values for each phase at the individual supply points are connected by a line.

III. Next Steps

Impact calculations of photovoltaic systems and usage of electric vehicles on the load profile of residential low voltage power grid must take into account of the facts unbalanced consumers three-phase loads along with popular and economic available single-phase rooftop photovoltaic systems with/without battery units and charging of electric vehicles. This means that we cannot count on a decrease or increase in load uniformly, but we must divide load into the individual phases and assume the installation of PV or EV on one of these phases. This step seems to be complicated due to the fact that smart meter customer data contains only the current effective value of voltage in a specific time interval and load for a certain time period (Fig.5).

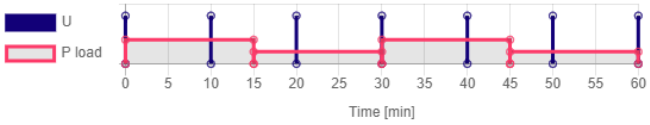


Fig. 5. Load profile and Voltage measurement timeline

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576.

References

[1] D. T. Nguyen, L. B. Le, "Optimal Bidding Strategy for Microgrids Considering Renewable Energy and Building Thermal Dynamics" IEEE Transactions on smart grid, Vol. 5, No. 4, 2014

[2] "Solcast API Toolkit", Solcast [Online], Available at https://solcast.com/solar-data-api/api-toolkit/

[3] "AS-6P30 Polycrystalline module specification", AmeriSolar [Online], Available at https://www.weamerisolar.eu/wp-content/uploads/2017/03/AS-6P30-Module-Specification.pdf

[4] F. Benavente, A. Lundblad, P. E. Campana, Y. Zhang, S. Cabrera, G. Lindbergh, "Photovoltaic/battery system sizing for rural electrification in Bolivia: Considering the suppressed demand effect", Applied Energy 235, 2019, pp. 519–528

[5] S. Ashok, "Optimised model for community-based hybrid energy system" Renewable Energy 32, 2007, pp. 1155–1164

[6] G. Gruosso, "Analysis of Impact of Electrical Vehicle charging on low voltage power grid" International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC), 2016

Multi-user adaptive interface for web-based collaborative virtual environments

[1]*Marián HUDÁK(4th year)*
Supervisor: [2]Branislav SOBOTA

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

[1]marian.hudak.2@tuke.sk, [2]branislav.sobota@tuke.sk

***Abstract*— This work introduces a Multi-user adaptive interface integrated into LIRKIS G-CVE web-based global collaborative virtual environments. In the last years, the rapid development of Extended Reality (XR) introduced novel technologies that differ in peripherals and inputs. Utilizing adaptive interfaces can be efficient to solve XR support for web-based CVEs. In this study, we present the adaptation of user peripherals to corresponding 3D interaction under WebXR standards. Our goal is to minimize the redundancy of user interfaces in collaborative virtual environments.**

***Keywords*— extended reality, web-based, user interaction, virtual collaboration, adaptive interface.**

I. INTRODUCTION

Globalization of web-based collaborative virtual environments (CVEs) gained potential for multipurpose use. However, global usage increases the number of users that access CVEs through different hardware platforms and operating systems [1]. This causes a variety of issues associated with multi-platform accessibility and cross-device support. In 2014, the JavaScript API standard called WebVR was released to provide access to commercial Virtual Reality (VR) devices via web browsers. The WebVR has positively affected the development of web-based 3D graphics libraries and frameworks for multipurpose VR usage. However, cross-platform development is extremely affected by the increasing novelty of technologies and devices. Therefore in 2017, the WebVR introduced the next standard called WebXR to open web access for Extended reality (XR) devices and interfaces. Extended reality includes the Virtual Reality (VR), Mixed Reality (MR), and Augmented Reality (AR) for the same purpose. Later in 2019, we implemented the LIRKIS G-CVE [2] web-based system that offers fully immersive and globally accessible CVEs under different platforms and web browsers. The general purpose of the LIRKIS G-CVE is to provide a testing web-platform for developing user XR interfaces and evaluating their usability and performance.

II. MULTI-USER ADAPTIVE INTERFACE

In general, globally accessible CVEs utilize a client-server architecture where the server replicates data between all connected clients in real time [3]. However, global CVEs are facing massive user access from different devices and platforms. Centralizing a wide range of user interfaces into the client-side causes frequent performance issues and high latencies during virtual collaboration. Therefore, most of the global CVEs provide only limited support of interaction and user devices. The introduction of XR technologies affected the whole development of CVEs [4]. The XR covers a wide range of devices and interfaces to fulfill users' experiences and interactions. Integrating XR support for global CVE purposes becomes the primary objective of researchers and experiments. Therefore, the multi-user adaptive interfaces (MUAI) were proposed to provide only the 3D interaction techniques that correspond to user devices and peripherals.

III. DESIGN AND IMPLEMENTATION

In the previous state, the LIRKIS G-CVE included redundant peripheral interfaces and scripts on the client-side. These negatively affected its performance and load balancing while running CVE. We aimed to enhance LIRKIS G-CVE with the MUAI interface for XR purposes by utilizing the WebXR standard. Our goal was to provide corresponding interfaces for client peripherals from the server-side and ensure their assignment to the proper 3D interaction techniques. Therefore, we proposed two MUAI interfaces (Fig.1) integrated on the client-side (MUAI-C) and on the server-side (MUAI-S). Both interfaces were implemented in JavaScript utilizing WebXR API.

The MUAI-C contains three components responsible for assigning proper 3D interaction techniques corresponding to the client's device and its peripherals. The first one represents the *WebXR component* which performs peripherals detection and requests the server to send the correct peripheral script. The second component called *3D Interaction Tools Assignment* compares the relation between CVE interactions and the peripheral features. The resulting comparison is processed by the *3D Interaction Technique Assignment* component, which attaches detected peripherals to proper 3D interaction techniques. Subsequently, the interaction techniques are executed as interaction commands and sent into the CVE.

The MUAI-S mediates all resources for clients and monitors their interaction commands in shared CVE. Each client informs MUAI-S of the resulting peripheral detection and chosen interaction technique. If some of the MUAI-C components work incorrectly, the MUAI-S obtains an error message from a client and requires him to reconnect. The error message contains issue description and the name of the troubleshooting MUAI-C component.

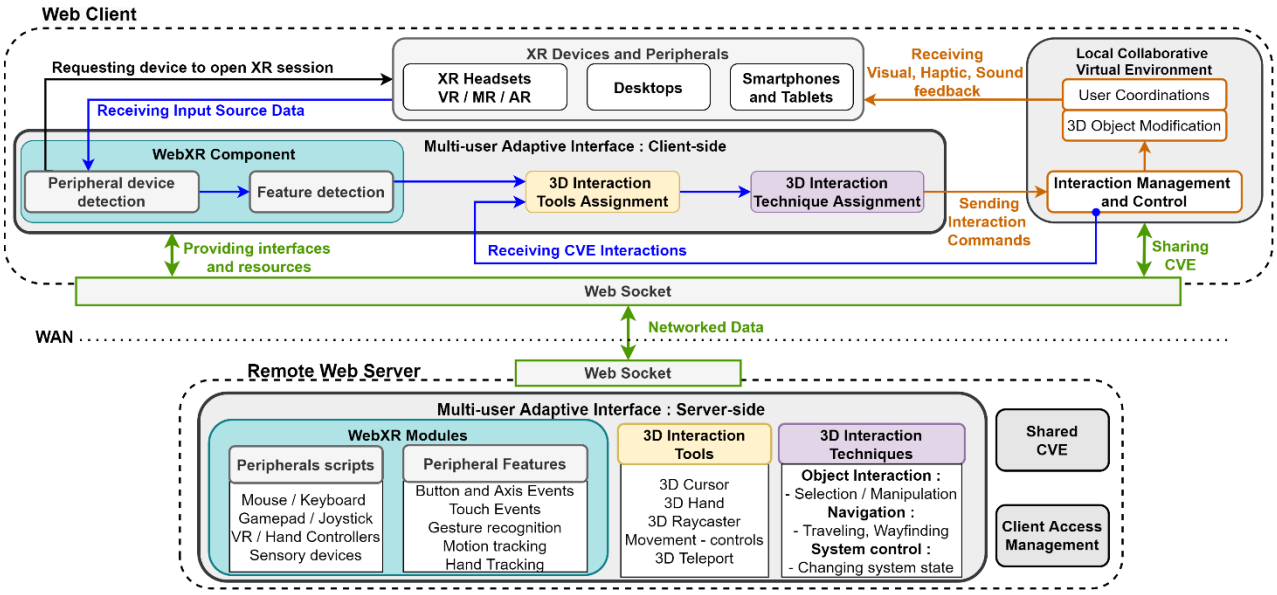


Fig. 1. The simplified model of LIRKIS G-CVE architecture with integrated Multi-user Adaptive Interfaces (MUAI-S and MUAI-C).

IV. Experimental Evaluation and Results

After integrating the MUAIs in the LIRKIS G-CVE, we performed several measurements to find out how the interface affects client latency when communicating to the server. Our experiments were carried out under Glitch cloud platform with Amazon Web Services (AWS) EC2 - M4 server instance. On the client-side, there were utilized Oculus Quest 1 and ASUS FX504 SERIES notebook (CPU Intel Core i5-8300H, 8.00GB RAM, NVIDIA GeForce GTX 1050 4GB). The experiment was held in two stages. In the first stage, the client obtained all peripheral scripts from the server and then ran the CVE while in the second stage the MUAIs were utilized. During both stages, the measurements started immediately after the client was connected to the server.

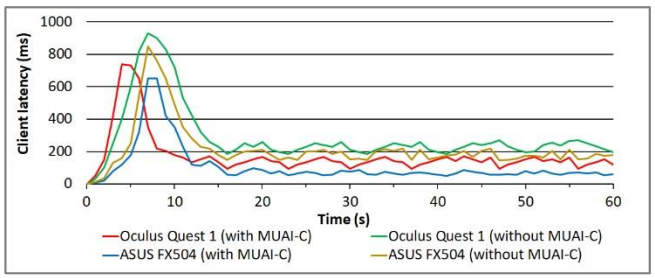


Fig. 2 Experimental results of measuring client latency with and without MUAI-C under Oculus Quest 1 and ASUS FX504.

After finalizing the experimental evaluation, it has shown, that utilization of both MUAIs positively affects the latencies on client-side. Each measurement was replicated 50 times. The results shown in Fig.2 describes average latency in milliseconds obtained within 60 seconds during the client communication to the server. In each of the measurements, it is visible that during the first 10 seconds the clients get resources and content from the server. When utilizing MUAI-C, the maximum latency on client-side was 720 ms (during the first 10 seconds) while without MUAI-C the latency increases in each case above 860 ms. After the latency stabilized, the lowest values were visible when utilizing MUAI-C with ASUS FX504 (49ms) and Oculus Quest 1 (62ms). In contrast, the highest latencies reflected on ASUS FX504 (200ms) and Oculus Quest 1 (290ms) when both MUAIs were deactivated. Each experiment resulted in observation that utilization of MUAIs positively affected the reliability of user interaction. Before deploying MUAIs, the interaction techniques were often unrelated to client devices. This caused a significant failure of interaction. Integrating MUAIs enhanced the client interaction by providing only interaction techniques that correspond with its peripherals.

V. Conclusion

In this paper, we have introduced the Multi-user adaptive interface for XR purposes. We demonstrated the whole process of adapting user inputs to 3D interaction techniques for controlling virtual environments. We assume that our contribution can be useful for the next research of web-based CVEs and user interfaces. Based on the experimental results, we confirm that reducing the number of client-side interfaces can increase its performance and reliability of user interaction.

Acknowledgment

This work has been supported by the APVV grant no. APVV-16-0202 "Enhancing cognition and motor rehabilitation using mixed reality" and by the KEGA grant No. 035TUKE-4/2019: „Virtual-reality technologies and handicapped people education “.

References

[1] R., Khadka, J. H.,Money, & A., Banic. Evaluation of scientific workflow effectiveness for a distributed multi-user multi-platform support system for collaborative visualization. In Proceedings of the Practice and Experience on Advanced Research Computing, 2018, pp. 1-8.

[2] M., Hudák, Š., Korečko, B., Sobota. LIRKIS Global Collaborative Virtual Environments: Current State and Utilization Perspective. In: *Open Computer Science,*2020, 11(1), pp.99-106.

[3] M., Hudák, Š., Korečko, B., Sobota. Enhancing team interaction and cross-platform access in web-based collaborative virtual environments. In: *2019 IEEE 15th International Scientific Conference on Informatics*, IEEE 2019, pp.171-176.

[4] M., Hudák, Š., Korečko, B., Sobota. Advanced User Interaction for Web-based Collaborative Virtual Reality. In: *2020 11th IEEE International Conference on Cognitive Infocommunications.* IEEE 2020, pp.343-348.

Utilization of battery energy storage systems in frequency regulation

[1]*Jakub URBANSKÝ(4th year)*
Supervisor: [2]Ľubomír BEŇA

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

[1]jakub.urbansky@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract* —Energy storage systems play a significant role in proper integration of renewable energy resources in maintaining reliable and modern power system. They can reduce power fluctuation, enhance electric system flexibility, and enables the storage and dispatching of the electricity produced by variable renewable energy sources. Battery energy system seems to be necessary addition for future of the power system. This paper deals with utilization of battery energy storage in power systems, specifically in frequency regulation.**

***Keywords*—frequency regulation, battery energy storage systems, energy storage, BESS.**

I. Introduction

The electric power system, as any part of the industry, is undergoing a constant development. In the recent years has tendency to incline more toward so called “green sources” or in other words renewable energy sources (RES), mainly from environmental reasons. The Renewable Energy Directive enact goal for European Union to achieve its 20% renewables target by year 2020 [1] and 27% by year 2030 [2]. Electric power systems need to preserve stable balance between production and consumption of electric energy. Penetration of RES such as, photovoltaics and wind energy has brought new issues for the stable operation of power system, due to their unpredictable electricity production character. Recent development in field of RES. mainly in photovoltaic, and emission free transport brought increased interest in energy storage. Due to unpredictable electricity generation from RES, energy storage plays important role in maintaining stable and secure operation of power grid, system flexibility and enables the storage of electricity in time of their increased production. On other hand, at time of insufficient production it can dispatch stored energy for their immediate usage. Energy storage can be utilized in power systems as stable source of ancillary services to provide reliable operation of power grid. My dissertation thesis will be focused on collaboration of renewable energy sources with battery energy storage systems (BESS), mainly on utilization of BESS in frequency regulation and self-sufficient usage of RES with battery storage collaboration.

II. Recent research

Our preliminary research is focused on investigating usage of RES and energy storage in field of power engineering. Serval usage scenarios were examined in past for example implementation of small wind turbine in simulation carried out in Matlab Simulink also usage of electric vehicle in household as power source and energy storage device, implementation of PV[3][4]. Since overhead power lines are still relevant topic even in the new age of power engineering focused on so called smart grids, we were examining also ampacity of overhead power lines[5][6].

III. Current research

Our current research is focused on utilization of energy storage systems in frequency regulation (FCR-N – frequency containment reserve for normal operation). There was created a model of BESS in simulation software Matlab Siumulink (see Fig. 1 and Fig. 2). Capacity of system is 1MWh and maximal output power is 1MW. Simulation step was 1s.

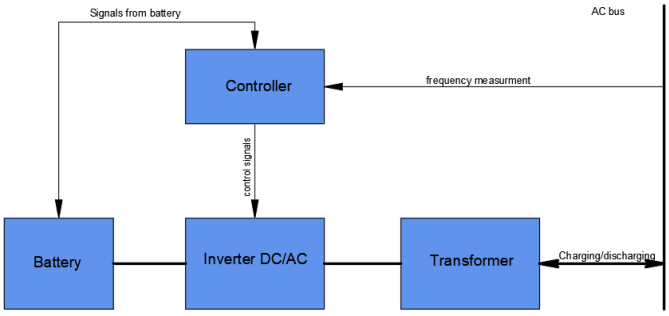


Fig. 1 Block diagram of the model

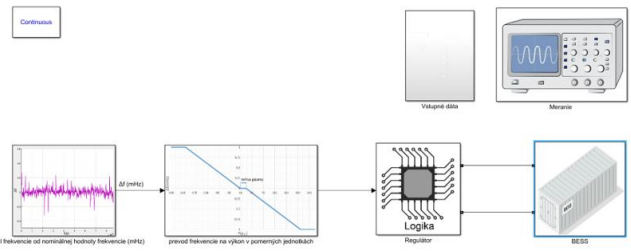


Fig. 2 Implemented model in Matlab Siumulink Software

Model consist of lithium ion batteries, controller unit, two-way DC/AC inverter, transformer (in simulation it is considered as reduced efficiency of the system).

In the model we are considering charging strategy of maintain state of charge of BESS around 50%. Energy in model is dispatched from and to BESS according to algorithm shown at Fig. 3.

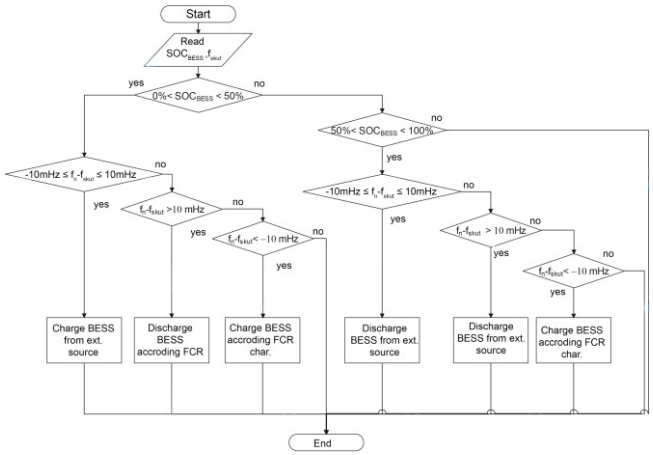


Fig. 3 Control algorithm for maintaining SOC by charging / discharging in the dead band

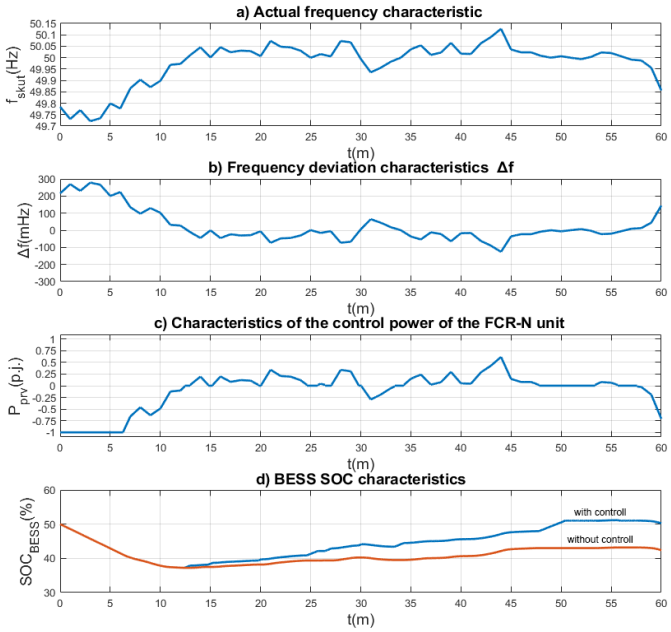


Fig. 4 Characteristics of a) actual frequency, b) frequency deviations, c) control power of PRV unit, d) charge level BESS (with and without SOC maintenance) for a time interval of 60m

At Fig. 4 a) it is possible to see actual characteristic of frequency. At Fig. 4 b) there is calculated frequency deviation from nominal frequency of 50Hz. At Fig. 4 c) there is characteristic of power output of controller according to frequency deviation based on FCR-N characteristic shown at Fig. 5. At Fig. 4 d) there is comparison of BESS SOC (state of charge) with and without control strategy. Characteristics at Fig. 4 b) c) d) are output of our model.

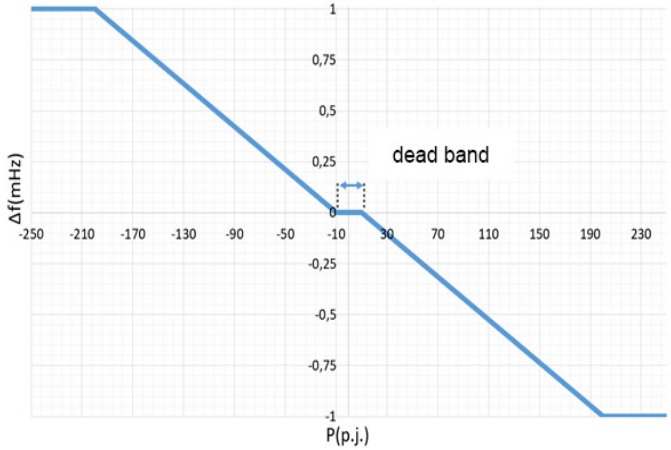


Fig. 5 FCR-N characteristics with dead band

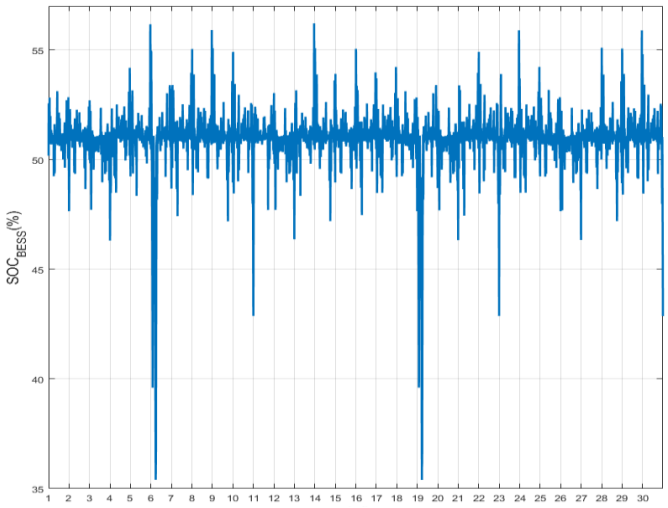


Fig. 6 Characteristics of the BESS SOC level in 30 days

We have simulated month consisted of 30days. The maximum SOC level of BESS for the simulated month was 56.208% and the minimum was 35.402%. The average SOC value for the simulated month was 50.937%.

We also estimated battery degradation level caused by charging and discharging shown shown in Table I.

We successfully developed working model of FCR-N with control strategy for maintain BESS SOC at level around 50%.

TABLE I
DEGRADATION OF BESS CAPACITY

SOC deviation(%)	Lower limit (%)t	Upper limit (%)	Cycles (-)	Capacity loss (%)
<2	0	2	44813	0,0265412
From 2 to 4	2	4	107	0,0001644
From 4 to 6	4	6	23	$8{,}4436.10^{-6}$

FUTURE RESEARCH

In our future research we want to estimate investment return of 1MW/1MWh BESS project used for FCR-N. We would like to also develop hybrid model of RES source with BESS collaboration using optimization software HOMER pro and examine optimization its features.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No.by APVV-19-0576

REFERENCES

[1] European Commission: Renewable energy directive[Online]. Available at: <https://ec.europa.eu/energy/en/topics/renewable-energy/renewable-energy-directive> Accessed on February 13, 2021

[2] European Commission: 2030 Energy Strategy [Online]. Available at: < https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2030-energy-strategy> Accessed on February 13, 2021

[3] J. Urbanský et al., "Implementation of Small Wind Turbine in Matlab Simulink, " In Poster 2019, pp. 1-4.

[4] J. Urbanský et al., "Elektromobil ako zdroj elektrickej energie v rodinnom dome," In Elektroenergetika, Vol 13 No 1 (2020)

[5] J. Urbanský et al. "Measurement of the influence of ambient conditions on the current carrying capacity of the 143-AL1 25-ST1A conductor," In Projektowanie, badania i eksploatacja, pp. 353-360. ISBN 978-83-66249-24-0

[6] J. Urbanský et al. "Determination of overhead power lines ampacity based on cigre 207 brochure," In Elektroenergetika 2019, Proceedings of the 10th International Scientific Symposium on Electrical Power Engineering, pp. 474-478. ISBN 978-83-66249-24

Available knowledge of 64x100 Gb/s DWDM system with different modulations

[1]*Norbert ZDRAVECKÝ (1st year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]norbert.zdravecky@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*— In this article we describe the topology and theory of DWDM system, the phenomenon affecting the transmission of the optical network and our goal of 64-channel DWDM system with center frequency of 193.1 THz, which will be designed in the Optisystem program. The theory of DPSK modulation is described. We will evaluate BER and Q factor parameters for 100 Gb/s systems with changing DPSK and DQPSK modulation.**

***Keywords*— BER, DPSK, DQPSK, DWDM, non-linear effects, Q-factor.**

I. Introduction

In DWDM system frequency spacing between adjacent channels should be less than or equal to 200 GHz [1]. This technique is used to make full use of huge bandwidth resources of low-loss standard single mode fiber. Here the optical signals of different frequencies are highly coupled and transmitted through a single fiber. The basic components of optical communication systems are a laser, optical fiber and receiver. DWDM (Dense Wavelength Division Multiplexing) systems transmit signals from multiple optical sources over a single fiber. DWDM systems must include some components to combine incoming signal. Multiplexer takes optical wavelengths from several fibers and converts them into a single beam [2]. Demultiplexers perform the function of separating the individual light components from the received beam. Demultiplexing is performed before light is detected by the photodetector, because photodetectors are essentially broadband devices that cannot selectively detect a single wavelength. Increasing transmission power leads to a distortion of the transmitted signal due to the nonlinear Kerr effects or interference between adjacent channels. These are the main limitations that set the upper limit of the maximum effective SNR (Signal-to-Noise Ratio) of the optical link and limit the performance of the whole system [3].

II. Comparison of DWDM Research

There are many research results about DPSK performance analysis and DWDM systems. Chandra and Bose show analytic results of the DPSK modulation in DWDM system with maximum ratio combining (MRC) in a single input and multiple output environment. The main scope of Chandra and Bose is analyzing single input and multiple output diversity effect on system performance. It does not solve the repetition scheme [4]. S. Vats and V. Kakar shows the performance of 64-channel 40 Gb/s DWDM system with frequency spacing 200 GHz. It was used RZ modulation and EDFA dopped amplifier with gain 10 dB. Conclusion from this work is that maximum Q factor and minimal BER is at launch power -2 dBm [5].

III. Modulation Methods

A key functionality of an optical system is therefore the modulation operation, it converts the high bit-rate electrical data signal into the optical domain. Ideal modulation is therefore equivalent to performing a frequency translation from the baseband to an optical carrier frequency [6, 7]. An optical transmission system of more than 40 Gb/s needs to be established to succeed the conventional 10 Gb/ss transmission system. However, if 40 Gb/s signals with narrow pulse width for 1 bit and wide frequency bands are transmitted through the existing fiber transmission system, the signals are affected by amplified spontaneous emission noises which occur and accumulate at the repeater amplifiers, as well as by chromatic dispersion or polarized mode dispersion at the erbium doped fiber amplifier (EDFA) or in the fibers. It is well known that the differential phase shift keying (DPSK) transmission does not require a reference phase and reduces the computational complexity and power consumption of the receiver at the cost of the transmission reliability. To improve the reliability of transmission under such scenario, we use a symbol repetition scheme to the DPSK modulation. However, the message symbol repetition with interleaving in conventional manner as in coherent modulation schemes may degrade the receiver performance because of the noise characteristic of the differential demodulation structure [8]. Fig. 1 shows DPSK modulation system and Table 1 shows parameters for DPSK modulation.

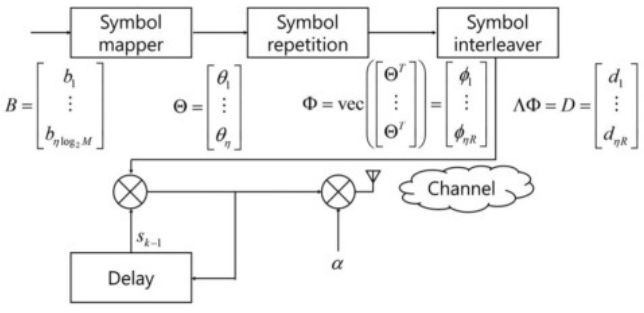


Fig. 1 Differential phase shift keying modulator system model

TABLE I PARAMETERS AND IDENTIFICATION

Parameters	Identification
η	the number of symbols
B	input bits sequence
b_k	kth input bit
Θ	presymbol sequence
Λ	permutation matrix for interleaving
Φ	repeated presymbol sequence
s_k	kth symbol
D	interleaved presymbol sequence

Differential QPSK (DQPSK) is a variant that is compatible with noncoherent receivers that mean receivers do not synchronize the demodulation oscillator with the modulation oscillator. Differential QPSK encodes data by producing a certain phase shift relative to the preceding symbol. By using the phase of the preceding symbol in this way, the demodulation circuitry analyzes the phase of a symbol using a reference that is common to the receiver and the transmitter. By using relative phase instead of absolute phase, DQPSK is not affected by a fixed phase offset introduced by lack of phase synchronization between transmitter and receiver; the fixed offset affects both symbols equally and is eliminated in the subtraction process. DQPSK is also robust against transmitter–receiver frequency discrepancies.

IV. Phenomena affecting transmission losses

Losses by nonlinear scattering occur especially at higher optical power densities in the optical fiber. Nonlinear phenomena cause the transition of optical power from one fiber to another, which propagates in the same or opposite direction. The performance dependence of the refractive index is responsible for the Kerr effect. Depending on the type of signal input, the Kerr effect is manifested by three effects such as SPM (Self-Phase Modulation), CPM (Cross-Phase Modulation) and FWM (Four-Wave Mixing). At high performance levels, Stimulated Brillouin-Scattering (SBS) and Stimulated Raman-Scattering (SRS) are manifested [9].

A. Raman Scattering

It is one of the nonlinear phenomena affecting the transmission via optical fiber. The classical description of the interaction of radiation with intramolecular motions is based on the consideration of the dependence of the electronic polarizability of α molecules on nuclear coordinates. The electron polarizability of the molecules can be calculated using equation (1):

$$\alpha(Q) = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 Q + \dots \quad (1)$$

Q is the temperature movement in the substance, which is equal to spontaneous Raman scattering. Induced dipole moment depends on the sum (anti-Stokes Raman) and differential (Stokes Raman) frequency. The subscript depends on the number of molecules in the macroscopic volume defined by the light field. [10]. The force can cause resonant oscillations. However, the radiation field consists of two waves - the exciting radiation wave and the Stokes wave. Under such conditions, there are chaotic intramolecular movements having a fluctuating character, they impose forced oscillations, the phases of which in individual molecules are determined by the phases of the components of the radiation field (phasing of molecular vibrations) in the macroscopic volume defined by the light field.

Stimulated Raman Scattering (SRS) - is a threshold phenomenon that occurs when the intensity of a light wave of frequency ω incident on a substance environment exceeds the threshold value $I \geq I_{thr}$. When this condition is met, the Stokes wave with frequency ω amplifies exponentially as it propagates through the environment. Assume its propagation in the positive direction of the z-axis, g denotes the gain coefficient. The exponential gain is calculated using equation (2):

$$I_S = I_{S0} \exp\left(g I_L z\right) \quad (2)$$

Stimulated Raman scattering leads to the transfer of energy from the doped wave to the constricted high-frequency and higher frequency waves through the transmission of the optical phonon in the transmission medium. The stimulated gain coefficients for different processes are determined by the properties of the optical phonons. The approach commonly used to describe the Raman interaction in optical fibers is to divide the problem into equations for doped waves and Stokes wave cascades [11].

B. Brillouin scattering

Stimulated Brillouin Scattering (SBS) is classified as a nonlinear process that allows the detection of optical components. It was first observed in 1964. The threshold value is calculated as for SRS. Exceeding the threshold results in the generation of a return Stokes wave that carries a higher input power. SBS limits channel performance in optical communication systems. It focuses on the generation of Stokes waves, the frequency of which is shifted downwards from the incident light by a value that depends on nonlinear media. The SBS phenomenon is similar to the SRS phenomenon. The first difference between SBS and SRS is that the Stokes wave in SRS propagates in both directions and in SBS it is only the reverse direction within single-mode fibers [12]. Stimulated Brillouin scattering (SBS) is well known for its excellent performance in phase conjugation, pulse compression, beam shaping, Brillouin-enhanced fourwave-mixing slow light and beam combination. For instance, SBS phase conjugate mirrors (SBS-PCMs) can be adopted to compensate a beam distortion caused by optical elements in the system, which would improve brightness and performance of high power lasers. SBS was demonstrated in a number of different media including liquids, gases, and fibers. To improve SBS performance, new media are exploited such as heavy fluorocarbon liquids, chalcogenides, photonic crystal fibers (PCF), on-chip waveguides, silicon, and diamond [13]. In Brillouin scattering a pump photon at a frequency ω_p produces an acoustic phonon at a frequency ω_Ω and a downshifted

V. Design of 64 channel DWDM system

The OptiSystem allows users to plan, test and simulate schematics of various kinds. For advanced modulations as DPSK and DQPSK is added a subsystem with this modulation. In the WDM transmitter, we set the frequency of 193.1 THz [16]. Frequency separation for the 100 Gb/s system, is 200 GHz. Another important parameter is the assigned power of 0 dBm, which we can change according to requirements. The channel part uses a loop that passes through the channel system only once. Inside are two amplifiers that are set to amplify a 10 dB signal. Used single mode optical fiber is 60 km long. The multiplexer is set to receive 64 channels and send a one complex signal to channel part. The demultipexor receives one signal as opposed to the multiplexer and transmits more signals according to the user's needs. In addition to the signal separation function, it provides frequency assignment parameters, which should be equal to the transmitting part, frequency bandwidth separation. Optical receivers are a suitable choice for receiving a signal from a demultiplexer. The last part is BER analyzer that examines and transparently renders the BER function. The Q factor is a function of the OSNR (Optical Signal to Noise Ratio) and represents the tolerance of the system in dB. BER and Q factor are close parameters. Equation (3) expresses the formula for calculating the BER and equation (4) expresses the formula for calculating the Q factor.

$$BER = \frac{1}{2} erfc\left(\frac{Q}{\sqrt{2}}\right) \approx \frac{1}{\sqrt{2\pi}Q} \exp\left(-\frac{Q^2}{2}\right) \quad (3)$$

$$Q = \frac{I_1 - I_0}{\sigma_1 + \sigma_0} \quad (4)$$

In these equations, I_1 is the mean value and σ_1 is the impediment of pulse 1, I_0 is the mean value and σ_0 is the deviation of pulse 0. These parameters create a comprehensive DWDM system for basic modulations. On Fig. 2 is shown DPSK subsystem simulation scheme in Optisystem. On Fig. 3 is shown DQPSK subsystem simulation scheme in Optisystem. On Fig. 4 is shown scheme of a 64x100 Gb/s system that will be simulate in Optisystem.

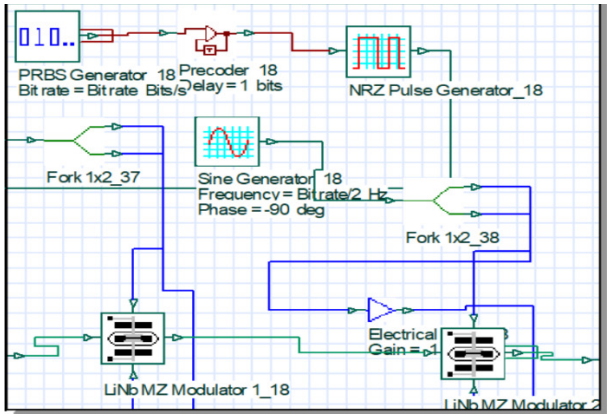


Fig. 2 DPSK modulation scheme in Optisystem

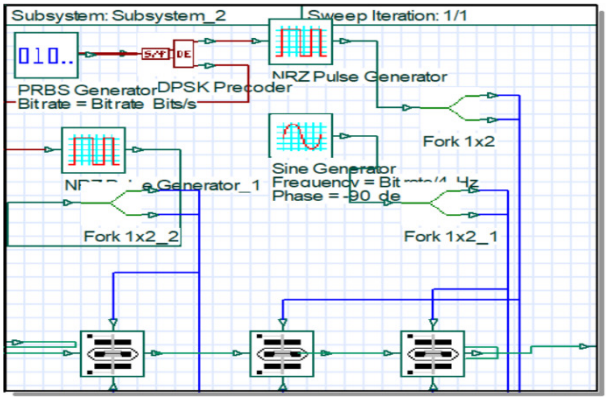


Fig. 3 DQPSK modulation scheme in Optisystem

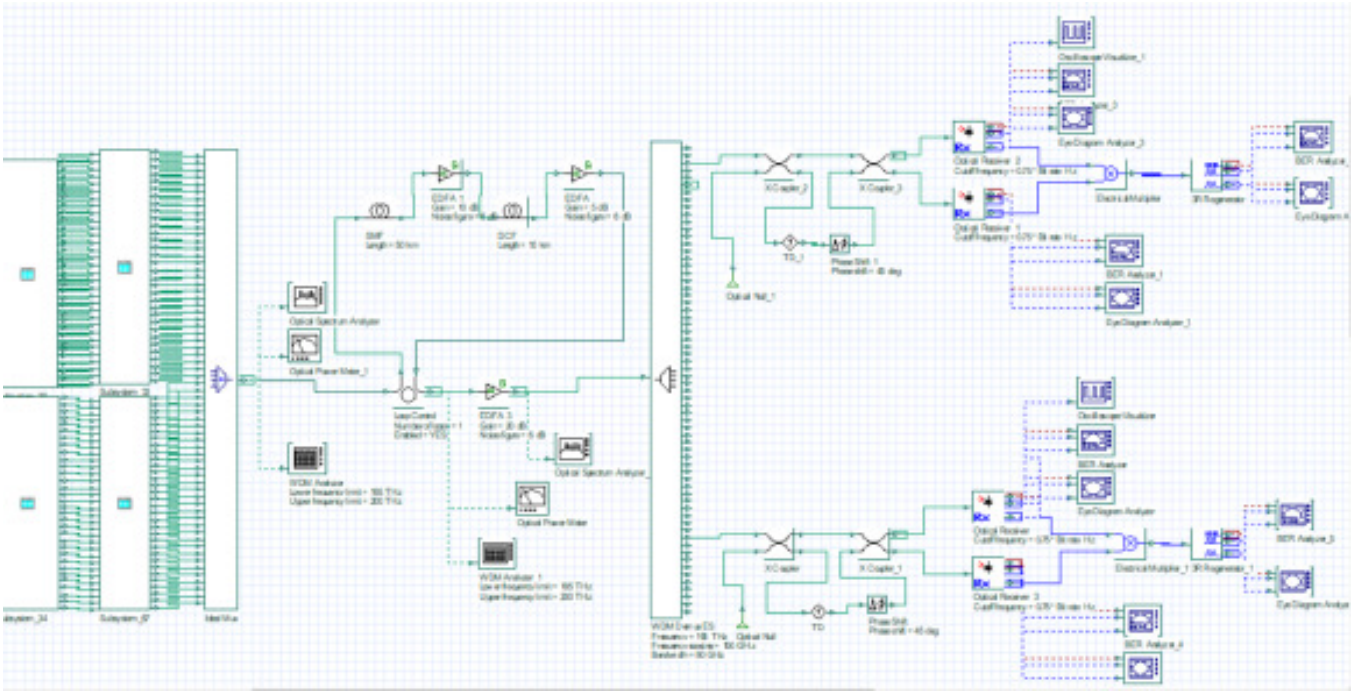


Fig. 4 General diagram of the 64-channel system from the OptiSystem program

VI. Future Work

Our future work is defined by dissertation thesis:

1) Design of high-capacity Ultra-DWDM systems for the investigation of nonlinear effects (SPM, XPM and FWM) for transmission speed above 40 Gb/s and 100 Gb/s in fully optical transmission networks.
2) Design of implementation of modified advanced optical modulation formats to reduce the impact of nonlinear phenomena for Ultra-DWDM optical transmission systems.
3) Optimization of optical link with the optical amplifier (EDFA) transmission paths in high-capacity Ultra-DWDM optical transmission systems.

Acknowledgment

This work was supported by research grant of FEEI TU Košice no. FEI-2021-71.

References

[1] G.694.1: Spectral grids for WDM applications: DWDM frequency grid. ITU-T, 2020. Available on: https://www.itu.int/rec/T-REC-G.694.1/en

[2] WINZER, Peter J. Energy-Efficient Optical Transport Capacity Scaling Through Spatial Multiplexing. IEEE Photonics Technology Letters. 2011, 23(13), 851 - 853. ISSN 1041-1135. Available on.: doi:10.1109/LPT.2011.2140103

[3] HUSZANÍK, Tomáš, Ján TURÁN and Ľuboš OVSENÍK. On the Impact of Fiber Nonlinear Effects on the CP-DQPSK Modulated Ultra-DWDM System. Acta Electrotechnica et Informatica. 2019, 19(4), 21-28. ISSN 1335-8243. Available on: doi:10.15546/aeei-2019-0026

[4] CHANDRA, Aniruddha and Chayanika BOSE. Series solutions for pi/4-DQPSK BER with MRC. International Journal of Electronics. 2011, 99(3), 391-416. Available on: doi:10.1080/00207217

[5] VATS, Saurabh and Vidur KAKAR. Performance Analysis of 64-Channel DWDM System Using EDFA. International Journal of Innovative Research in Computer Science & Technology. 2015, 3(2), 79-82. ISSN 2347-5552.

[6] HUSZANÍK, Tomáš, Ján TURÁN and Ľuboš OVSENÍK, "Performance Analysis of Optical Modulation Formats for 10 Gbit/s DWDM System," Carpathian Journal of Electronic and Computer Engineering, vol. 10, no. 2, 2017, pp. 3-8. ISSN 1844 – 9689.

[7] KAUR A., S. DEWRA. "Comparative Analysis of Different Modulation Techniques in Coherent Optical Communication System," International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, issue 8, p. 7193 – 7200, August 2015. ISSN: 2320-9801.

[8] HAN, Sungmin, Jaeseok LEE, Taesoon KWON and Ji-Woong CHOI. Performance analysis on DPSK modulation using symbol repetition and interleaving. Communication systems. 2018, 31(11), 1-16. Available on: doi:10.1002/dac.3589

[9] SENIOR, John M. Optical Fiber Communications: Principles and Practice. 3rd Edition. London: Pearson, 2008. ISBN 978-0-13032-681-2

[10] SUR, Ujjal K. Surface-enhanced Raman scattering (SERS) spectroscopy: a versatile spectroscopic and analytical technique used in nanoscience and nanotechnology. 2013, 1(2). Available on: doi:10.12989/anr.2013.1.2.111.

[11] AGRAWAL, Govind P. Optical Fiber Communications: Optics and Photonics. 5th Edition. Cambridge: Academic Press, 2012. ISBN 978-0-12397-023-7

[12] MOHAMAD, N.R, A.A WAKIRAN and M.F ISWANDI. Development of Optical Wireless Audio SystemUsing Infrared Light Communications. Journal of Electronics and Communication Engineering. 2013, 8(6), 65-69. ISSN 2278-2834

[13] PANT, Ravi and Christopher G. POULTON. On-chip stimulated Brillouin scattering. 2011, 19(9), 8285-8290. Available on: doi:10.1364/OE.19.008285

[14] RAKICH, Peter T., Charles REINKE and Rayan CAMACHO. Giant Enhancement of Stimulated Brillouin Scattering in the Subwavelength Limit. Physical Review. 2012, 2(1), 1-13. ISSN 2278-2834. Dostupné z: doi:doi.org/10.1103/PhysRevX.2.011008

[15] KAZOVSKY L. G., W. T. SHAW, D. GUTIERREZ, N. CHENG, S. W. WONG, „Next-Generation Optical Access Networks", Journal of Lightwave Technology, vol. 25, no. 11, 2017. DOI: 10.1109/JLT.2007.907748.

[16] HUSZANÍK T., J. TURÁN, Ľ. OVSENÍK, "Experimental Simulation of Coherent 40 Gbps 2-DPSK DWDM Long-haul Fiber Optical System with Counter-directional EDFA," Acta Technica Napocensis, vol. 59, no. 2, 2018, pp. 1 – 4.

[17] ALI R., M. S. ALI, T. MIR, B. SHABIR, and U. F. LILLA, "Analytical Review of Advance Optical Modulation Formats," Electr. Electron. Eng. An Int. J., vol. 4, no. 1, pp. 35–43, 2015.

[18] CHHILAR R., J. KHURANA, and S. GANDHI, "Modulation Formats in Optical Communication System," vol. 13, no. July, pp. 110–115, 2011.

An overview of Autonomous Mobile Robot research and multi cooperation

[1]*Stanislav ALEXOVIČ (1st year)*
Supervisor: [2]Milan LACKO

[1,2]Department of Electrical Engineering and Mechatronics, FEI TU of Košice, Slovak Republic

[1]stanislav.alexovic@tuke.sk, [2]milan.lacko@tuke.sk

***Abstract*—This paper is intended as an overview of the research area of Autonomous Mobile Robot (AMR). It describes in first part basic design and usage of AMR with related problems and already used solutions and strategies to solve these problems. Second part is dedicated to cooperation and communication among multiple AMRs, fleet architecture design and description of algorithms which could be used to solve path planning problems.**

***Keywords*—AMR (Autonomous Mobile Robot), AGV (Autonomous Guided Vehicle), Fleet.**

I. Introduction

With increased automatization in industry, Autonomous Mobile Robots are more and more used to perform various tasks in the logistics segment of factories. Usually they are assigned to a task which contains moving from point A to point B and some basic manipulation. Usually they are loaded with an operator or other arm robot then they are moving to destination stations where they are unloaded or if they contain some automatic load/unload device they are capable to do this task fully automatically [1].

Another area of usage are warehouses. This is a logical and straightforward application for AMRs because of their nature to carry stuff but the main problem here is that AMR have to be capable of automatically loading and unloading items from and to shelfs. This could sound like an easy task but in reality it is a difficult operation. AMR has to perfectly know its environment and also items which are going to operate. But this could be partially solved with adjusting the environment for AMR. It means that AMR will operate only with items which are suitable for it.

And there is one new area of usage for these robots. It is in a civil environment. There are a lot of opportunities to carry some stuff but cooperation in this environment is really hard because it is very variable, always changing, and interacting with people. But again, we could find some subareas where AMR would fit. E.g. for cleaning surfaces in hospitals [2] or offices during night to avoid interaction with people. Or a really new concept of disinfection of surfaces with germicidal UV lamps “Fig. 1”.

II. AMR (Autonomous Mobile Robot)

AMR or Autonomous Mobile Robot is the successor of AGV (Autonomous Guided Vehicle). AMR are considered as more intelligent robots which are capable of better navigation in its environment. AGVs are navigated with magnetic tape on the floor. It is closer to the railway system. While AMR is capable of navigation without this floor tape. Basically, you define a goal station and AMR should be able to find a path there and even avoid collisions with people or other objects which it meets during movement [3].

Fig. 1. AMR with germicide addon.

Eyes for AGV are only sensors for magnetic tape. And navigation is to follow tape until it reaches goal station which has to be somehow marked on the floor. But AMR needs more sophisticated eyes and more intelligent navigation. There are two main components which can provide this functionality. 3D cameras and lidars. Both measure distance to the surrounding environment [2].

A. Navigation

Navigation is the process of finding a path from point A to point B while tracking the actual position. To know the actual position is an essential requirement of navigation. This could be achieved in various ways. Maybe the simplest solution is to use GPS but for many AMRs this is not an option. Because AMR usually operates under a roof where there is no GPS signal.

Another option is to use a known position of some signal beacon. E.g. Wi-Fi access points. If AMR is able to receive signals from at least three beacons then relative position to these beacons could be determined. This process is called trilateration. For these solutions AMR has to be able to determine beacon straight distance. This could be achieved with measurement of time which is required for signals to travel from beacon to AMR. One distance defines the perimeter of possible positions. If AMR is able to collect three of these distances then position is the intersection of all perimeters.

Fig. 2. Vizualization of lidar lasers.

There is also one relatively simple solution for position tracking. It is called wheel odometry [5]. And it means that AMR is precisely tracking movement of wheels. If AMR is able to measure linear velocity and also angular velocity then it should be able to determine position in time. But this technique has one big disadvantage. It is precision of measurement. In the real world measurement is never ideal and error is increasing in time. So longer this technique is used then a bigger inaccuracy in position is received.

But this could be a complementary technique to something else. And it is measuring distance to static objects in an AMR environment. Lidar and 3D cameras are used for this. But these sensors are weak if the observable environment is not changing. Imagine a long narrow corridor and AMR moving through. Even though AMR is moving all observable elements have always the same distance. So for calculating position relative to static objects some inequality is required. And exactly in this case wheel odometry could be helpful to determine position.

Algorithm which is able to combine these two approaches is called AMCL (Adaptive Monte Carlo Localization). Is probabilistic localization system which is based on particle filter to track position against known map [4].

B. SLAM

All previous navigation approaches used some known entities. Position of beacons, starting position or map. But what to do if none of this is known? This area of research is called SLAM (Simultaneous Localization and Mapping).

SLAM is "is the computational problem of constructing or updating a map of an unknown environment while simultaneously keeping track of an agent's location within it." Input for SLAM are sensors. Mainly 3D cameras or lidars. If AMR can "see" its environment, then SLAM algorithms can determine position relative to seen objects. If the robot starts to move, then seen objects are observed from different angles. Therefore, the SLAM algorithm can determine the new position of AMR. It means that movement can be tracked.

But what to do if objects in the environment are also moving? Then it is hard to determine what is moving. If it is an AMR or observed object. For this reason, SLAM must use also other sensors for detecting its own movement. It is the accelerometers, gyroscopes and magnetometer. All together form Inertial Measurement Unit (IMU). Which can provide orientation, position and velocity. So, if AMR with SLAM can detect its own pose change and also surrounding changes due to movement then SLAM can provide exact pose inside the observed environment.

To successfully operate AMR in some environment both of attitudes must be used. First stage is to use SLAM to map new unknown environment and create map from it "Fig. 3". This could be done manually with controlled movement of AMR or automatically where AMR is roaming around to discover a new environment. And when a map is created the operator can perform a second stage where he/she specifies some points in the known map where AMR should go to perform some action. It is also possible to update map during the second stage if the observed environment is changing [6].

III. Fleet

This was a brief introduction into AMR navigation. But if somebody wants to use more AMRs simultaneously a new area of problems arises. How to do navigation, operations and management of more than one AMR? It is important to note that more AMRs does not mean that the fleet system will be N times more effective than one AMR while N is count of AMRs. E.g. fleet can contain ten AMRs and if nine of them are blocked by one then the effectiveness of the whole fleet equals to only one. Therefore, management of the fleet needs to be designed in a robust way to use the whole potential of multiple AMRs together [7].
The first question which needs to be answered is whether a centralized, decentralized "Fig. 4" or distributed system for fleet management will be used. Both have their pros and cons.

A. Centralized

This approach consists of building one powerful master unit which can control each AMR. But it would be impractical if the master would have to control each step of AMR. But usually the master sends the whole path which is followed by AMR. From this point of view the master needs to solve optimization problems of generating proper paths to achieve its goal. Because this approach knows the whole system it can find optimal solutions but even though this approach has some drawbacks [14].

- High network traffic - All AMRs communicate with master therefore it needs to be able to handle traffic from all AMRs simultaneously.
- Failure in master means failure of whole system.
- Lack of scalability - If the master system is scaled for a certain amount of AMRs then adding some more could result in overload of the master system.

B. Decentalized

Another approach is to leave organization of AMRs to themself. This will lead to faster responses of AMRs because they must solve their goals be themselves. To use this approach could be also beneficial to avoid problems mentioned in the centralized approach. But a decentralized system is weaker to find optimal paths because of deadlocks [4].

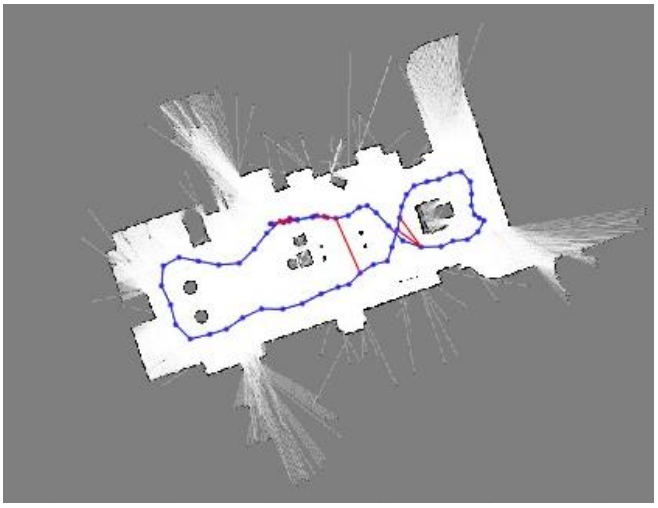


Fig. 3. Occupancy grid created by lidar SLAM.

C. Combined

It is usually a good approach to combine various approaches to get best results. In this case a centralized system could be used to plan global paths and goals. But reacting to environment events could be left for specific AMR. Collision avoidance is a good example. AMR could change the planned path to avoid obstacles and only notify the master system about this event. [9]

IV. Communication

Communication for AMR is essential functionality. It means that one AMR must be able to connect to another or to master a system depending on fleet architecture. Because of the nature of AMR communication must be wireless. For these purposes, a lot of wireless protocols could be used. E.g. Wi-Fi, Bluetooth, ZigBee. For main communication with master systems Wi-Fi is often used because of its widespread usage. But for communication between two AMR or direct communication with some other devices as charges, Bluetooth is more suitable [10].

Other layer of communication transport. Transport layer is responsible for sending and receiving data from and to AMR. Wi-Fi physical layer is often used with TCP/IP protocol. This protocol is heavily used for computers local and wide network. Therefore, AMR could be easily connected, and they behave in network as another computer. This is an advantage because an AMR logic program could be used as tools which are also available for computers [11].

Last layer of communication is the application layer. This layer describes the logic of communication. And it depends on the use case and requirements of the fleet. One popular pattern is publish/subscribe. It consists of publishers, subscribers and one broker. The major advantage of this pattern is that entities are decoupled. It means they do not know each other. Which support scalability and robustness of whole implementation. This pattern works in this way. Publishers only publish data. Usually the state of some component. (velocity, position). Subscriber subscribes for some specific data. It means that it receives data from the publisher. Communication entity between publisher and subscriber is called topic. Topic has its own name and on this Topic name publisher can publish something and subscribers are able to receive data from Topic. But they do not have to

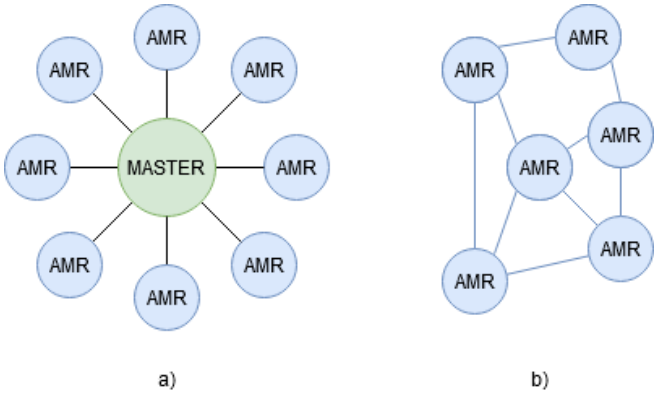


Fig. 4. a) centralized, b) decentralized arhitecture of fleet management system.

know who the publisher is and who is the subscriber. All this is managed by broker which is joining publisher to subscriber over topic [12].

V. Path planning Algorithms

It is an intelligent process of finding a path from point A to point B while avoiding obstacles which are placed in the environment. The Resulting path could be defined as a curve or set of small segments connected. This quite popular research subject which has been studied since 1970. And resulting algorithms are used in robotics but also in the game industry [13]. All these algorithms are solving optimization tasks of finding the best path. The best path could not be considered as only the shortest but also time and cost factor needs to be taken in account. Sometimes the shortest path could lead through areas which will result in other costs for the robot or whole fleet of robots.

According [13] path planning methods could be categorized according to environmental conditions as static and dynamic. Static is defined as an environment which is not changing in time. And dynamic is an environment which Global planner with local planer for obstacle avoidance.
could be changed in time. It means that it could contain elements which are moving and are considered as obstacles. For that reason, path planning algorithms are divided into two categories. Global and local. Global algorithms are meant for static environments or at least for static parts of the environment. And their job is to find the optimal path through this environment. Whilst local planning algorithms are used for finding paths to avoid obstacles which were not considered during global planning. Local planners use robot sensors to detect obstacles and avoid them. While a global planner is not able to see moving objects in the environment "Fig. 5".

Path planning algorithms are in nature optimization problem algorithms. There is a wide range of heuristics methods which could be used to solve this problem. Such as Particle Swarm Optimization (PSO), Artificial Neural Networks (ANN), Fuzzy logic (FL) and Genetic Algorithm (GA) [15].

VI. Path planning in fleet

Planning paths simultaneously for multiple AMRs is a complex problem. Because one AMR to another means obstacle. The main requirement of the fleet is to generate

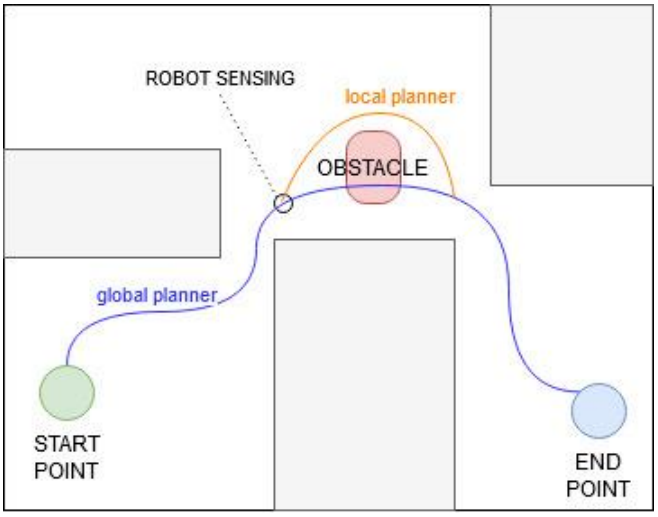


Fig. 5. Global planner with local planer for obstacle avoidance.

paths which will not collide, and each AMR time is used as effectively as possible. Another problem besides collision avoidance is deadlock. Meaning that group of AMRs is waiting for each other and none of them is moving. The problem is known as multiple robot path coordination problem and leads researchers to finding advanced approaches to how to solve it. In the author leaves bigger autonomy for specific robots. But they share the same network of paths. Each robot must acquire free paths to be able to use them. Fleet server is responsible for keeping track of acquired path segments. This approach should prevent collisions. But deadlocks and overall effectiveness is questionable.

Another approach was to plan paths in a static environment and later adjust velocities of robots to prevent collisions [17]. This approach was later tuned with prioritization in a way that the robot got priority level and therefore it was preferred during conflicting movement [18].

Prioritization is a useful strategy also in decentralized fleet architecture [19]. For decentralized system Leader-Follower algorithm could be used [20]. In this scenario one robot is the leader and knows the global goal and another robot or robots are only followers. Their responsibility is only to keep track of the leader and follow him. Leader supplements work on the master system for a group of robots.

VII. Conclusion

This paper was written as an overview of the Autonomous Mobile Robot research area. Intention was to collect information about usage of AMRs, related problems and methods used to solve these problems. Another part was about cooperation among multiple AMRs and identifying problems and possible solutions for these problems.
Intention of the author is to continue in research of multiple robot path coordination problem and try to design algorithm for effective usage of each AMR.

VIII. Acknowlegment

This work was supported by Slovak Research and Development Agency under project APVV-18-0436.

References

[1] Vancea, A. & Orha, Ioan. (2019). A survey in the design and control of automated guided vehicle systems. Carpathian Journal of Electronic and Computer Engineering. 12. 41-49. 10.2478/cjece-2019-0016.

[2] J. Bačik, F. Ďurovský, M. Biroš, K. Kyslan, D. Perduková and S. Padmanaban, "Pathfinder–Development of Automated Guided Vehicle for Hospital Logistics," in IEEE Access, vol. 5, pp. 26892-26900, 2017, doi: 10.1109/ACCESS.2017.2767899.

[3] Jianqiang Jia, Weidong Chen and Yugeng Xi, "Design and implementation of an open autonomous mobile robot system," *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, New Orleans, LA, USA, 2004, pp. 1726-1731 Vol.2, doi: 10.1109/ROBOT.2004.1308073.

[4] S. Zhou, G. Cheng, Q. Meng, H. Lin, Z. Du and F. Wang, "Development of multi-sensor information fusion and AGV navigation system," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 2043-2046, doi: 10.1109/ITNEC48623.2020.9084687.

[5] S. Quan and J. Chen, "AGV Localization Based on Odometry and LiDAR," 2019 2nd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM), Shanghai, China, 2019, pp. 483-486, doi: 10.1109/WCMEIM48965.2019.00102.

[6] Y. Chen, Y. Wu and H. Xing, "A complete solution for AGV SLAM integrated with navigation in modern warehouse environment," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 6418-6423, doi: 10.1109/CAC.2017.8243934.

[7] S. Hoshino and H. Seki, "Multi-robot coordination for jams in congested systems," Robotics and Autonomous Systems, 2013.

[8] S. LaValle and S. Hutchinson, "Optimal motion planning for multiple robots having independent goals," IEEE Transactions on Robotics and Automation, vol. 14, no. 6, pp. 912–925, 1998.

[9] V. Digani, L. Sabattini, C. Secchi, and C. Fantuzzi, "Hierarchical Traffic Control for Partially Decentralized Coordination of Multi AGV Systems in Industrial Environments," in 2014 IEEE International Conference on Robotics and Automation (ICRA), May 2014, pp. 6144–6149.

[10] S. H. Robinson, "Living with the challenges to functional safety in the industrial Internet of Things," Living in the Internet of Things (IoT 2019), London, UK, 2019, pp. 1-6, doi: 10.1049/cp.2019.0160.

[11] H. Chao and Z. Yufeng, "The Wireless Communication Method between Industrial robot and AGV based on SQLServer," 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 2019, pp. 481-484, doi: 10.1109/EITCE47263.2019.9095115.

[12] J. Deogirikar and A. Vidhate, "An improved publish-subscribe method in application layer protocol for IoT," 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, 2017, pp. 1070-1075, doi: 10.1109/SmartTechCon.2017.8358534.

[13] P. M. Bhushan Mahajan, "Literature review on path planning in dynamic environment," International Journal of Computer Science and Network, vol. 2, no. 1, pp. 115–118, 2013.

[14] H. L. A. Ramirez-Serrano and G. C. Pettinaro, "Mobile robot localization in quasi-dynamic environments," Industrial Robot: An International Journal, vol. 35, pp. 246–258, 2008.

[15] D. R. P. Prases K. Mohanty, "Controlling the motion of an autonomous mobile robot using various techniques: a review," Journal of Advance Mechanical Engineering, vol. 1, pp. 24– 39, 2013.

[16] V. Roberge, M. Tarbouchi, and G. Labonte, "Comparison of parallel genetic algorithm and particle swarm optimization for real-time uav path planning," Industrial Informatics, IEEE

[17] K. Kant and S. W. Zucker, "Toward efficient trajectory planning: The path-velocity decomposition," Int. J. Robot. Res., vol. 5, no. 3, pp. 72–89, 1986.

[18] M. Erdmann and T. Lozano-Pérez, "On multiple moving objects," in Proc. IEEE Int. Conf. on Robotics and Automation, San-Francisco, CA, 1986, pp. 1419–1424.

[19] R. Alami, F. Robert, F. F. Ingrand, and S. Suzuki, "Multi-robot cooperation through incremental plan-merging," in IEEE Int. Conf. on Robotics and Automation, Nagoya, Japan, 1995, pp. 2573–2578.

[20] J. P. Desai, J. Ostrowski, and V. Kumar, "Controlling formations of multiple mobile robots," in Proc. IEEE Int. Conf. on Robotics and Automation, 1998, pp. 2864–2869.

Magnetic properties of amorphous ferromagnetic glass coated microwires

[1]*Simeon SAMUHEL (1st year)*
Supervisor: [2]Jozef ONUFER

[1,2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]simeon.samuhel@tuke.sk, [2]jozef.onufer@tuke.sk

***Abstract*— Glass-coated microwires are perfect materials for theoretical study as well as for practical applications. Amorphous ferromagnetic glass-coated Fe-based microwires exhibit bistable behavior. These microwires are characterized by a rectangular hysteresis loop. In this work, some magnetic properties of amorphous glass-coated microwires are described. The study of dynamics of a single domain wall, its velocity, nucleation of reversed domains and critical fields in amorphous ferromagnetic Fe-based microwires is presented.**

***Keywords*— amorphous ferromagnetic microwire, domain wall, bistable behaviors.**

I. INTRODUCTION

Amorphous magnetic glass-coated microwires have attracted considerable attention because of their interesting magnetic properties and their use in sensor devices and other applications [1], [2], [3]. The domain structure of microwires with positive magnetostriction consists of one large domain in the central part of the metal core, which is covered by a domain structure with a radial direction of magnetization (Figure 1). In addition, small closure domains (domain structures) appear at the wire end to reduce stray fields. The magnetization reversal occurs by one large Barkhausen jump [4], [5]. The hysteresis loop of such a microwire is perfectly rectangular and the magnetization can have only two values $\pm M_s$ (where M_s is the saturation magnetization). Switching between two magnetization values occurs at a critical, so-called switching magnetic field H_{sw}. The switching field is highly dependent on external parameters such as temperature [6], mechanical stress [7] etc.

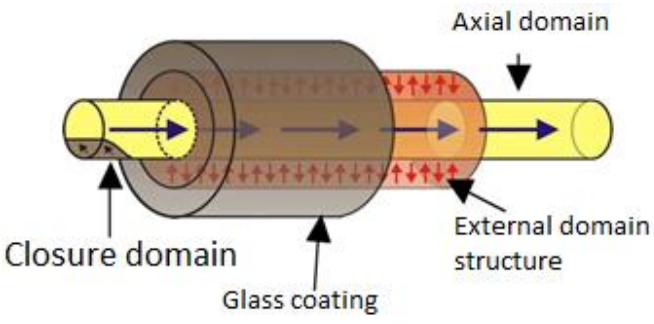


Figure 1 Schematic domain structure of amorphous glass-coated microwire with positive magnetostriction [8].

Magnetization process runs through large Barkhausen jump of a single domain wall along entire microwire. This process provides possibility to study a single domain wall propagation. As a result of their complex anisotropy distribution, the glass coated microwires exhibit very fast domain wall propagation, sometimes faster than sound speed [9]. There are two processes that are usually the subjects of study. The first one is the process of free domain wall release. This occurs by depinning a domain wall from the closure domain structure at the wire end [10] or by nucleation of a reversed domain in regions far from the microwire ends [11]. The second topic consists in measurements of domain wall velocity as a function of an external magnetic field ($v(H)$ dependences) [12].

A magnetic glass-coated microwire is a composite material that consists of a metallic nucleus of 0.5–30 μm in diameter and glass coating of thickness of 1–20 μm. Glass-coated magnetic microwires are prepared by so called Taylor-Ulitovsky technique [13]. In this way, the Taylor-Ulitovsky method is the unique technique allowing fabrication of microwires by rapid melt quenching. The presence of a flexible, thin, biocompatible, and insulating glass-coating allows to enhance the corrosive resistance. And therefore makes these microwires suitable for novel applications including biomedicine, electronic article surveillance, non-destructive monitoring external stimuli (stresses, temperature) in smart composites or construction health monitoring [14], [15], [16].

II. THEORY

In general, microwires are divided into three groups according to magnetostriction. The orientation of the circular shell magnetization domain structure is usually considered for Co-based microwires with negative magnetostriction. Fe-based microwires with positive magnetostriction have essentially an outer domain envelope with a radial orientation surrounding one axially magnetized domain (figure 1). The last group are microwires with low magnetostriction. Their domain structure is characterized by a large number of circular domains below the surface of the metal core and the structure of the axial domain in the center of the wire.

The main emphasis will be on microwires with positive magnetostriction, which show spontaneous magnetic bistability. The popular study of remagnetization processes is associated with bistable microwires. Remagnetization processes includes the depinning or nucleation and movement of one or more domain walls.

Damping mechanisms, such as eddy current damping and spin relaxation damping, have a great influence on the remagnetization process during the movement of a single domain wall.

The magnetic properties and dynamics of a single domain wall have been studied for last two decades [17]. These topics are still very important, both in terms of basic research and practical applications. This is due to the complexity of understanding the dynamics of domain walls, which depend on internal and external mechanical stresses distribution [18], magnetostriction and heat treatment. The complexity lies in the fact that these factors are often interdependent. Thus, the actual value of magnetostriction usually differs from the composition of the microwires and depends on the internal mechanical stress induced in process of fabrication [18]. Clarifying the dynamics of the domain wall could allow a deeper understanding of nature of the extremely fast propagation of domain walls in magnetic microwires, and consequently control the domain wall motion more effectively.

The domain wall propagation

Understanding and describing the dynamics of domain walls (movement of the boundary between magnetic domains), became an important topic with a view to the dynamics of remagnetization processes in microwires and their use in practice followed. Thanks to such a study, microwires have found application in many ways, including a wide range of magnetic logic devices and nanoscale data storage devices [19].

Domain wall displacement plays a significant role in the remagnetization process [20]. This requires the essential mechanisms of nucleation of reverse domains and of propagation of domain walls upon their enlargement.

The motion of the domain wall induced by the applied field in the ferromagnetic microwire is similar to the oscillation of a mechanical oscillator system, with the body oscillating under the influence of an external force in a viscous medium [21], [19]. By analogy, the equation for domain wall propagation is:

$$m\frac{d^2x}{dt^2} + \beta\frac{dx}{dt} + \alpha_t x = b\mu_0 M_s H \tag{1}$$

where b = 2 for 180 ° domain wall, m is the effective wall weight, x is the position of the domain wall, β is the damping coefficient of the domain wall, α_t is the coefficient proportional to the back pressure on the domain wall [21], M_s is the saturation magnetization and H is the applied magnetic field. Suppose a domain wall that propagates at a constant velocity. From this linear dependence we can derive the relationship between the domain wall velocity and the applied magnetic field H [22]:

$$v = S(H - H_0) \tag{2}$$

where the coefficient of proportionality S is the mobility of the domain wall:

$$S = \frac{b\mu_0 M_s}{\beta} \tag{3}$$

and H_0 is the critical field of domain wall propagation:

$$H_0 = \frac{\alpha x}{b\mu_0 M_s} \tag{4}$$

under which the spread of the domain wall is not possible.

An important task for the study of domain wall dynamics is to determine the damping coefficient of the domain wall β. Several terms contribute to such a damping. The first mechanism is the effect of eddy currents, which prevent the domain wall propagation. For the eddy current dumping coefficient β_{ec} we can write the equation [23]:

$$\beta_{ec} = \frac{{\mu_0}^2 {M_s}^2}{\rho} R_{ad}\left(\ln\frac{R}{R_{ad}} + \frac{8}{\pi^2}\right)\frac{2R_{ad}}{l} \tag{5}$$

where ρ is the resistivity (for our samples $\rho = 1.3.10^{-6}$ Ωm), R_{ad} is the radius of the axial domain, l is the length of the wall in the axial direction and R is the radius of the metal core. For our samples $R_{ad} \approx R \ll l$.

The contribution of damping from the relaxation of magnetic moments, the so-called the spin relaxation damping coefficient β_{sr} [23], is given by:

$$\beta_{sr} = \frac{2M_s\alpha}{|\gamma|}\sqrt{\frac{3\lambda_s\sigma}{2A_{ex}}} \tag{6}$$

where α is the Gilbert damping coefficient, γ is the gyromagnetic ratio, A_{ex} is the exchange interaction constant, λ_S is the saturation magnetostriction, and σ is the tensile stress.

III. EXPERIMENTAL

The Sixtus-Tonks experiment

The most widely used method of examining domain wall velocity was introduced by Sixtus and Tonks in 1932 [19][24]. The method is based on measuring the time for which the domain wall travels a defined distance between two places (pick-up coils).

A very simple experiment consisting of a primary coil (solenoid), two magnetizing coils and two pick-up coils. Primary coil creates a homogeneous field in which the domain wall propagates. The magnetizing coils are located at the ends of the wire. Then the domain wall propagates from particular end of the wire. A signal from two pick-up coils is measured by the oscilloscope. After passing single domain wall through the pick-up coils, two sharp pulses (peaks) appear on the oscilloscope (positive and negative if the coils are properly connected). The speed of the domain wall is then given by the ratio of the distance between the pick-up coils L and the time between the peaks t:

$$v = \frac{L}{t} \tag{7}$$

In the pick-up coil, the magnetization changes over time at a specific location of the sample, due to the movement of the domain wall. As a result, an induced voltage is generated at the ends of the coil due to a change in magnetization. The pick-up coils are very short and the distance L between them is large.

In this method, it is necessary to rule out the possibility that one wall is successively running along the microwire and at the same time another opposite domain is formed in the volume of this microwire (more domain walls). The movement of multiple walls can affect the result of the speed measurement. It can deal with this problem in different ways. One way is to use a system of multiple coils, where one pair of coils is used to measure the speed and the other to check the existence of other walls. On the other hands we can use one pair of coils for measurement and other coils are used to prevent the movement of another domain wall. Thus, if we know that only one domain wall is propagating, the dependence of the velocity on the applied field $v(H)$ can be measured. In our experiments we used two blocking coils at the ends of the microwire, which allowed us to influence the direction of the domain wall propagation. In

the middle between blocking coils is a pick-up coil to measure individual points of the hysteresis loop.

Unidirectional effect

For a given part of microwire (the part of the wire between the pick-up coils in the Sixtus - Tonks experiment), it is possible to measure four different domain wall velocities [12], [25]. For example, say that the wall moves from the end A to the end B of the wire, then it is possible to distinguish two types of wall (head – to - head or tail – to - tail) depending on the direction of the magnetic field. If the letters A, B indicate the direction of movement of the domain wall, index 1 means the head-to-head wall type and index 2 the tail-to-tail wall type respectively, then the so-called unidirectional effect (Figure 2.) fulfil the condition [25]

$v_{AB1}(H) \approx v_{BA2}(H)$ and $v_{AB2}(H) \approx v_{BA1}(H)$.

In other words, the speed of the wall depends on the orientation of the magnetization to which the microwire is magnetized by the propagation of the wall. The physical origin of this phenomenon is unknown.

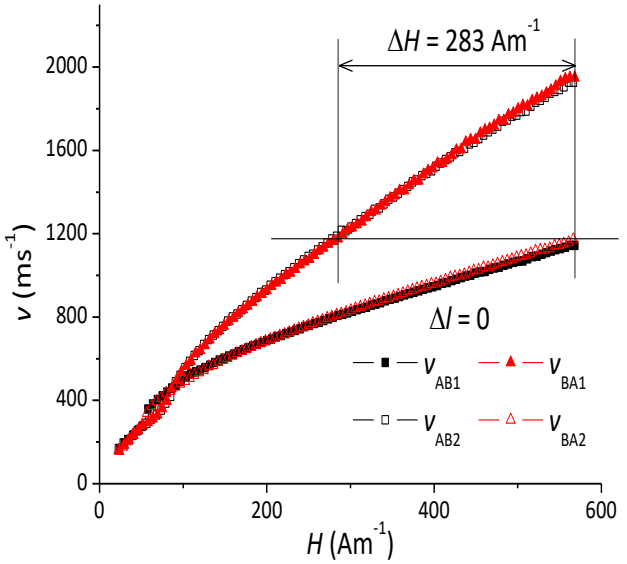


Figure 2 Domain wall velocity with unidirectional effect (FeSiB) [25].

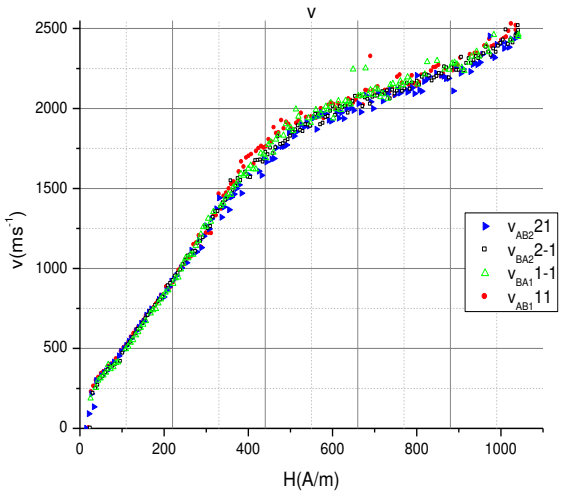


Figure 3 Four different velocities of a domain wall moving without unidirectional effect($Fe_{42,5}Co_{35}Si_{7,5}B_{15}$).

The unidirectional effect does not occur in every type of microwire (Figure 3). Using this experimental method, a unidirectional effect was found only for one type of Fe-based microwires [12], [25].

Distribution of critical fields

Another experiment was to measure the distribution of local critical fields [26], [27]. Local critical fields were measured using the following procedure. First, the wire was axially saturated using the negative magnetic field H_{max} generated in the solenoid. Then, a constant small positive H_{So} field in the solenoid was turned on. The H_{So} field was lower than the lowest critical field for magnetization change, but high enough for the existing free domain wall to move along the entire microwire. When a free domain wall was produced, the magnetization reversal in the pickup coil was detected. Subsequently, the final information on the magnetic state of the wire in the pick-up coil was obtained by means of an integrator. In the next step, the pick-up coil was shifted by 0.5 mm along microwire and the measurement was repeated until the distribution of critical fields along the entire microwire was measured (Figure 4). An amorphous glass-coated amorphous ferromagnetic microwire $Fe_{49,6}Ni_{27,9}Si_{7,5}B_{15}$ produced by the Taylor-Ulitovsky method with a metal core diameter of 15 μm was used.

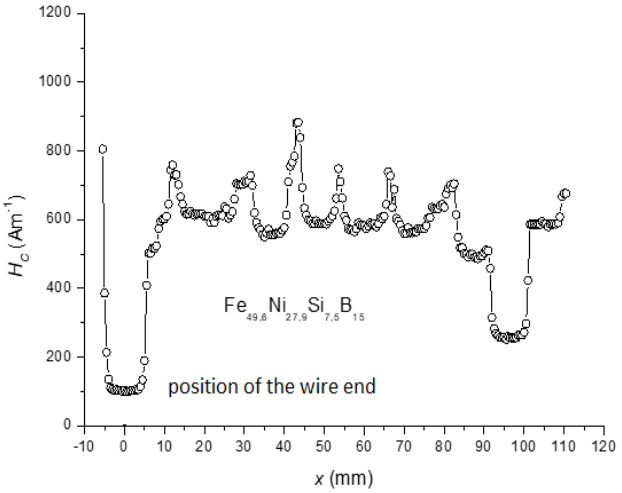


Figure 4 Distribution of local critical (nucleation) fields along the microwire (x - distance between the wire end and the center of the magnetized coil) [26].

Closure domain structure

The information about the closure domain structure by measurements of hysteresis loops could be obtained. Therefore, the following experiment was proposed. The scheme of the apparatus is shown in Figure 5. The set-up consists of a sliding coil placed inside a solenoid which generates a homogeneous magnetic field. The slide coil could be shift along the entire microwire.

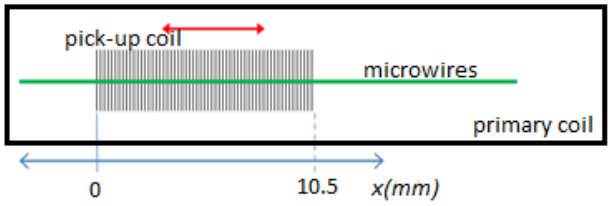


Figure 5 Experimental set-up. (x - distance between the wire end and the left end of the pick-up coil)

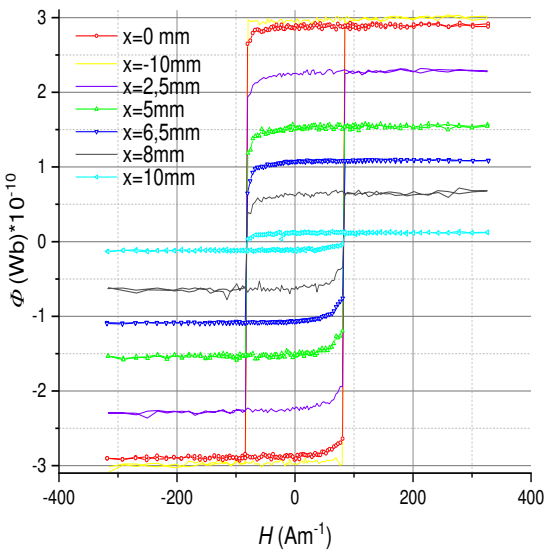


Figure 6 Measurement of hysteresis loops on different parts of the sample.

The rectangular shape of the hysteresis loops confirms the bistable behavior of the microwires with a positive magnetostriction constant. Axial loops were measured at different distances of the end of the sample from the end of the sensing coil. As can be seen, the measured results are as expected, with greater distance from the sensing coil the inductive flux decreases (Figure 6). This experiment could be helpful to study closure domain structure.

IV. CONCLUSION

The research results presented in this paper show that the microwires are still interesting both from a theoretical and an application point of view. The main emphasis was on microwires with positive magnetostriction, which show spontaneous bistability. The velocities of one domain wall were measured using a simple Sixtus-Tonks experiment. A unidirectional effect was observed when single domain wall moved in a different direction of magnetization, into which the microwire is magnetized. The process of release of a single domain wall from the closure domain structure at the microwire was studied.

REFERENCES

[1] ZHUKOVA, Valentina, et al. Electronic Surveillance and Security Applications of Magnetic Microwires. 2021.

[2] ZHUKOV, A.; IPATOV, M.; ZHUKOVA, V. Advances in giant magnetoimpedance of materials. In: *Handbook of Magnetic Materials*. Elsevier, 2015. p. 139-236.

[3] ZHUKOV, A., et al. Magnetic properties and magnetocaloric effect in Heusler-type glass-coated NiMnGa microwires. *Journal of alloys and compounds*, 2013, 575: 73-79.

[4] MOHRI, K., et al. Large Barkhausen and Matteucci effects in FeCoSiB, FeCrSiB, and FeNiSiB amorphous wires. *IEEE transactions on magnetics*, 1990, 26.5: 1789-1791.

[5] VÁZQUEZ, M.; CHEN, D.-X. The magnetization reversal process in amorphous wires. *IEEE transactions on magnetics*, 1995, 31.2: 1229-1238.

[6] ZHANG, Jianxing, et al. Giant high temperature superelasticity in Ni53Mn24Ga21Co1Cu1 microwires. *Intermetallics*, 2020, 122: 106799.

[7] SABOL, Rudolf, et al. Mechanical stress dependence of the switching field in amorphous microwires. *IEEE Transactions on Magnetics*, 2015, 51.1: 1-4.

[8] RICHTER, K., et al. Domain wall dynamics in thin magnetic wires under the influence of transversal magnetic field. *IEEE Transactions on Magnetics*, 2010, 46.2: 210-212.

[9] VARGA, R., et al. Supersonic domain wall in magnetic microwires. *Physical Review B*, 2007, 76.13: 132406.

[10] ZHUKOV, Arcady, et al. Studies of the defects influence on magnetic properties of glass-coated microwires. *IEEE Transactions on Magnetics*, 2014, 50.11: 1-4.

[11] ONUFER, J.; ZIMAN, J.; KLADIVOVÁ, M. Dynamics of closure domain structure in bistable ferromagnetic microwire. *Journal of magnetism and magnetic materials*, 2013, 344: 148-151.

[12] ONUFER, Jozef, et al. The influence of annealing on domain wall propagation in bistable amorphous microwire with unidirectional effect. *Physica B: Condensed Matter*, 2018, 540: 58-64.

[13] BARANOV, Serghei A.; LARIN, Vladimir S.; TORCUNOV, Alexander V. Technology, preparation and properties of the cast glass-coated magnetic microwires. Crystals, 2017, 7.6: 136.

[14] KOZEJOVA, D., et al. Biomedical applications of glass-coated microwires. *Journal of Magnetism and Magnetic Materials*, 2019, 470: 2-5.

[15] ZHAO, Y. J., et al. A self-sensing microwire/epoxy composite optimized by dual interfaces and periodical structural integrity. *Composites Part B: Engineering*, 2020, 182: 107606.

[16] CORTE-LEON, Paula, et al. Magnetic microwires with unique combination of magnetic properties suitable for various magnetic sensor applications. *Sensors*, 2020, 20.24: 7203.

[17] QIN, Faxiang; PENG, Hua-Xin. Ferromagnetic microwires enabled multifunctional composite materials. *Progress in Materials Science*, 2013, 58.2: 183-259.

[18] CHIZHIK, A., et al. Influence of combined mechanical stress on magnetic structure in magnetic microwires. *Journal of Magnetism and Magnetic Materials*, 2020, 513: 166974.

[19] ALLWOOD, Dan A., et al. Magnetic domain-wall logic. *Science*, 2005, 309.5741: 1688-1692.

[20] BECK, F.; RIGUE, J. N.; CARARA, M. The profile of the domain walls in amorphous glass-covered microwires. *Journal of Magnetism and Magnetic Materials*, 2017, 435: 21-25.

[21] VARGA, Rastislav, et al. Anomalous effects in the domain-wall dynamics in magnetic microwires. *physica status solidi (a)*, 2011, 208.3: 509-514.

[22] VARGA R.: Doménové steny a ich dynamika. Košice: Prírodovedecká fakulta UPJŠ, 2014. 199s. ISBN 978-80-8152-173-7

[23] DEL REAL, R. P., et al. Eddy current damping of planar domain wall in bistable amorphous wires. *Applied physics letters*, 1993, 63.25: 3518-3520.

[24] PONOMAREV, Boris Konstantinovich. Start field fluctuations of amorphous Fe_5Co_70Si_10B_15 alloys. *Fizika Tverdogo Tela*, 1984, 26.10: 2974-2979.

[25] ONUFER, J.; ZIMAN, J.; KLADIVOVÁ, M. Unidirectional effect in domain wall propagation observed in bistable glass-coated microwire. *Journal of Magnetism and Magnetic Materials*, 2015, 396: 313-317.

[26] ONUFER, J., et al. The Study of Closure Domain Structure Dynamics in Bistable Microwires Using the Technique of Three-Level Field Pulses. *IEEE Transactions on Magnetics*, 2018, 55.1: 1-6.

[27] ZIMAN, J.; ONUFER, J.; KLADIVOVÁ, M. DC magnetization processes in bistable glass-coated ferromagnetic microwires. *Journal of magnetism and magnetic materials*, 2011, 323.23: 3098-

Design of Battery Tester

[1]*Juraj BIĽANSKÝ (2nd year),*
Supervisor: [2]*Milan LACKO*

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

[1]juraj.bilansky@tuke.sk, [2]milan.lacko@tuke.sk

***Abstract*—The article describes the design of a device for cyclic testing of battery cells. It describes the basic concept of the whole device. The concept is divided into colored blocks. Besides, the components for the construction of the device are selected in the article and the resulting diagram of the power part of the battery tester is attached. Based on this scheme, the device shown in the last figure was built.**

***Keywords*—Battery, tester, design, power, mosfet, MCU**

I. Introduction

In recent years, batteries have become a very important part of our lives. They provide power for the portable electronic devices that we use everyday. For example, we can mention mobile phones, laptops, watches, electric cars and much more. With increasing technological advancements in portable electronic devices, the market for affordable batteries is also growing. Many different types of batteries are available on the market today, such as nickel-cadmium (NiCd), nickel-metal hydride (NiMh) or lithium-ion (Li-ion). The latter Li-ion batteries make up the majority of the market. They are used in mobiles, laptops and many other industries. These batteries are also used in electric vehicles. Batteries used in electric vehicles are called traction batteries. Traction batteries consist of a large number of small battery cells which, by suitable series–parallel connection, form an electric vehicle battery with the required properties such as output voltage, current and capacity. Final battery quality, then, defines the basic characteristics of the vehicle such as power, range and weight. From the point of view of cyclic testing of batteries, this process is energy, time and financially demanding (the battery pack is expensive, in the case of an electric car it is also heavy and has a high capacity) therefore, we will simplify testing and we will test the basic building blocks of these batteries, namely battery cells itself [1], [2], [3].

II. Battery Tester Conception

When choosing the concept, we focused on the modular design of testers, as we want to create several devices that will be connected via CAN bus to the main device, which will have a programmed HMI interface for setting tests and various charging and discharging profiles. The result is the following concept in Fig. 1. The concept is divided by colors:

- Blue - Non-power parts of the device (control and communication)
- Red - Power parts of the device where a higher current is expected
- Yellow - device control (in our case we assume digital control using MCU)
- Green - Battery

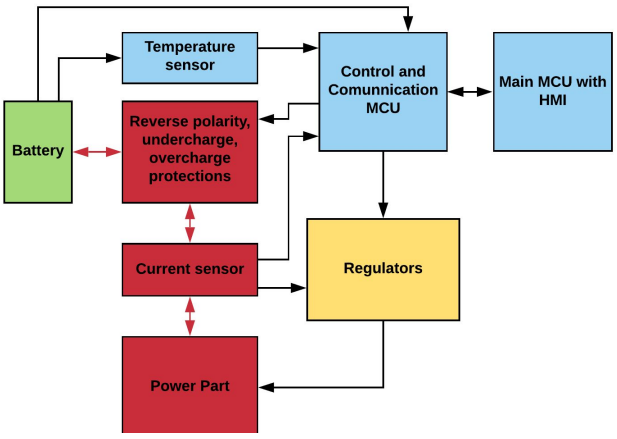


Fig. 1. Conception of battery tester [2]

III. Design of Individual Parts

Devices is able to discharge batteries with the following parameters:

- All types of Li-Ion batteries
- CC/CV charging with max current 5A and max voltage 4.2V
- Maximal discharging current 20A

A. Control and Comunnication MCU

The NUCLEO-F303K8 module with a DSP core was selected as the control and communication processor. This MCU supports all functions necessary for the controling of the power part, such as [4]:

- Core frequency up to 72MHz
- Up to 16 PWM channels
- Up to 9 ADC channels and 3 DAC
- SPI, I2C, USART, CAN

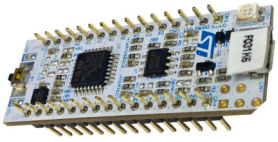


Fig. 2. NUCLEO-F303K8[4]

In addition to the communication with the master MCU, this processor performs a control activity, i.e. the regulation block is performed software in the processor.

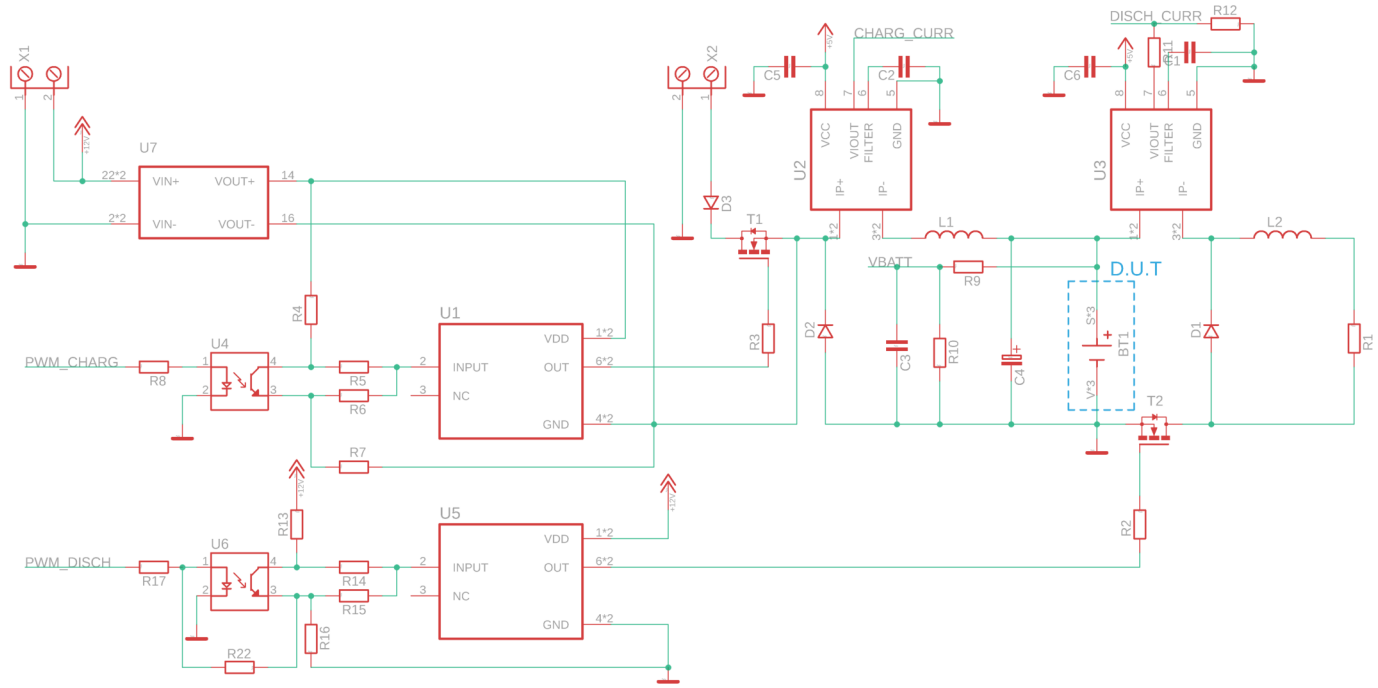


Fig. 3. Power part of schematic

B. Power Part

As a consequence of the different desired accuracy of current sensors for charging and discharging processes, the battery tester (Fig.3) consists of two step-down dc/dc converters

- Charging circuit – The circuit consists of the components X2, D3, T1, U2, L1, BT1 and D2. The 12V voltage from the power supply is changing using T1, L1 and D2 according to the PWM duty cycle. The control of the charging process is in previous article [3]. Besides power components, the charging circuit also contains the MOSFET driver U1, the isolated power supply U7 and the optocoupler U4.
- Discharging circuit – The circuit consists of the components BT1, U3, L2, R1, D1 and T2. The role of the power supply plays the battery BT1. Energy is consumed by all components but primarily by the 100 W power resistor. Besides playing the role of a load, the power resistor is responsible for short current protection since its value was chosen to constrain the maximum value of the discharging current to 25 A.

.

Fig. 4. Final device mounted to power supply

IV. Conclusion and Future Work

The article describes the design of a device that is used for cyclic testing of battery cells. It is a modular device design that will be connected via a CAN bus to a master MCU with HMI interface. The device is currently in the same state as in Fig. 4. A larger number of such modules (40 and more) is expected, which will test selected battery cells with a large number of cycles for 24/7.

Future work on the device will be its programming and tuning of the program to control the whole power part according to requirements, ie charging with constant current/voltage and discharging with a constant current or variable current, which will change over time according to our specified profile. Besides, stable protection protocols have to be programmed into the device to protect the cells during the charging and discharging processes. Last part that needs to be programmed is the battery analysis during test cycles when important data will be stored on the SD card and then evaluated. Thanks to this data accurate models can be created.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under Contract No. APVV-18-0436

References

[1] J. Bilansky and M. Lacko, “Design and simulation of cyclic battery tester,” *Power Electronics and Drives*, vol. 5, no. 1, pp. 229–241, 2020.

[2] J. Biľansky and M. Lacko, “Návrh koncepcie nabíjacej a vybíjacej Časti testéra li-ion batériových Článkov,” *Electrical Engineering and Informatics*, vol. 11, pp. 385–389, 2020.

[3] J. Biľanský and M. Lacko, “Overview of the battery types and their testing,” *SCYR 2020 - Nonconference Proceedings of Young Researchers*, pp. 143–146, 2020.

[4] STMicroelectronics, *NUCLEO-F303K8*, 2021. [Online]. Available: https://www.st.com/en/evaluation-tools/nucleo-f303k8.html

[5] A. MicroSystems, *ACS713*, 2021. [Online]. Available: http://www.allegromicro.com/~/Media/Files/Datasheets/ACS713-Datasheet.ashx

The Test Bench for High-Precision Actuators Testing

[1]Jozef IVAN (2nd year)
Supervisor: [2]František Ďurovský

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

[1]jozef.ivan@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*— This paper presents requirements for building the functional dynamometer. In the last section, the developed experimental system for testing high precision actuators where the load cycle is defined by a spreadsheet is presented. The experimental system disposes of a user interface for full control over measurement which facilitates the process of measurement.**

***Keywords*—Dynamometer, experimental system, variable load**

I. Introduction

In modern companies, there is a constant request for shortening the period of the processes connected with the production, and every delay is costly. Also, in the stage of development and testing, it's desired to test and evaluate every possible issue that can encounter in the future process. In such cases, it's suitable to imitate a specific type of load which will be driven by an electric machine in operating conditions. As it would be very difficult and costly to equip the laboratory with all the equipment to simulate all kinds of mechanical load, simulating with a dynamometer is a satisfactory alternative. Such a type of simulation is commonly known as emulation. In general, researchers use electrical drives to emulate different mechanisms, such as combustion engines [1], gas engines[2], wind turbines [3], electric vehicles[4], and more[4]. In this paper, the experimental dynamometer system for high precision actuator testing will be presented.

II. Dynamometer system

The dynamometer is a device that can deliver a particular steady-state torque-speed relationship and is often equipped with devices for measuring the position and speed of the shaft. Mostly we use the second electric motor as a dynamometer, which is coupled to a tested motor via a shaft. The proposed system is shown in Fig.1, where Load Machine (LM) represents dynamometer and the Drive Under Test (DUT) represents tested machine. T_e is the applied electrical torque of the drive machine, and T_L is the applied load torque of the load machine. To achieve the proper function of a system the actual position θ or actual angular speed ω of the shaft has to be measured.

III. Requirements on the experimental system

To construct a functional test bench, we have to consider multiple necessary components, such as a load machine, power converter which provides energy to the LM, equipment for measuring and gathering data from the testing system, construction of the test bench, and different types of sensors depending on measured variables.

The power and torque level of LM should be selected higher or equal to the power level of the DUT so it can't overload LM. If this is not the case, it is necessary to constrain the torque operation range of the drives. The LM should be capable to operate in all four quadrants so it can recuperate energy when operating in generator mode. Also, it is suitable to power both drives from the same DC-bus as it allows the transfer of energy between both machines. SMPM motor is often selected as LM because it has high dynamics and can be exposed to a short-term overload.

To achieve the best possible dynamic response with LM it is desirable to operate it in the torque control loop. The control of LM depends on the DUT actual torque value which can be obtained from an estimator or observer by monitoring DUT current. Also, the torque sensor attached to the shaft can be used. DUT mostly uses a speed or position control loop depending on the final implementation. If it's possible, it's beneficial to feed DUT with the same converter as in the final implementation so the whole technology can be tested. This way we can also avoid implementation of errors into control structure which can occur during code rewriting between laboratory and final technology.

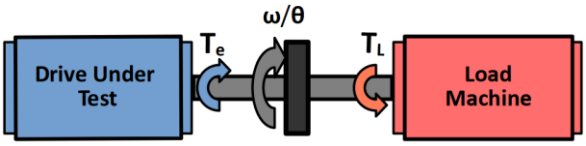


Fig.1 Basic structure of dynamometer

IV. Developed Experimental system

A subset of the test bench for testing high precision actuators is shown in *Fig.2* The whole system is designed to measure the performance of DUT when operating under a defined load cycle. The main concern is to monitor the temperature rise of the tested actuator to verify correct drive dimensioning. The other measured variables are torque, speed, and position of the shaft. Measured torque is used in a closed-loop torque control system of the LM, with a sampling period of 800 µs. The measured speed obtained from the DUT encoder is used in the closed-loop speed control structure of the tested machine.

Currently, the B&R 3,64 kW SMPM with 11,6 Nm nominal torque is used as LM. It is fed by B&R 8V1180 9 kW ACOPOS

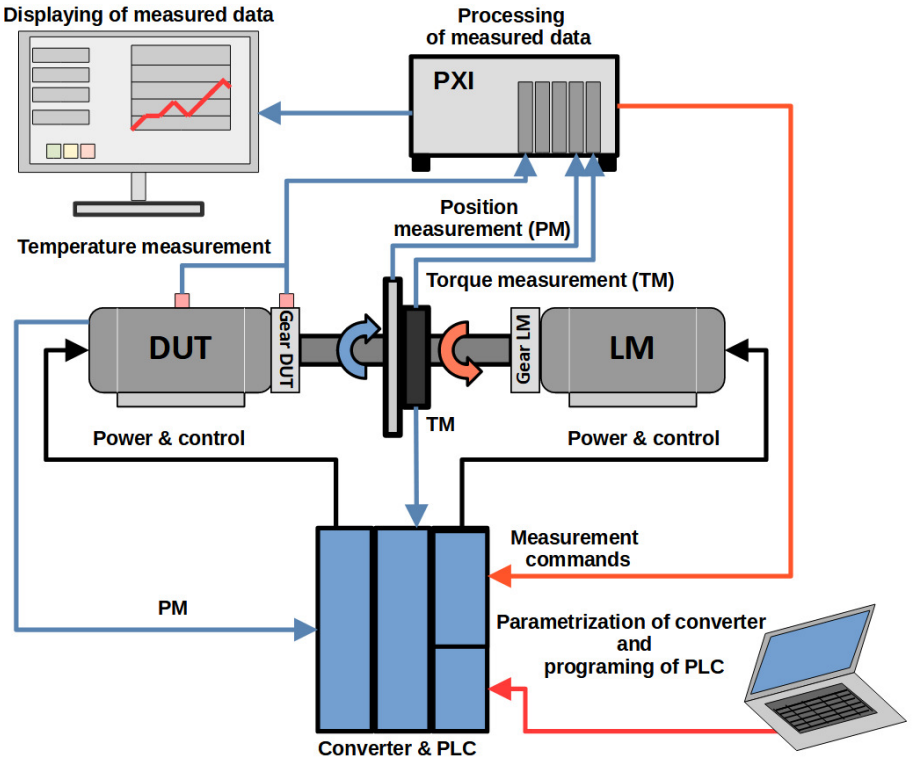


Fig.2 Experimental system

type converter. The measurement system disposes of multiple ACOPOS converters on different power levels for powering DUT. The LM is coupled with planetary gear with a 1/50 ratio. The actual torque is measured with Magtrol torque flange sensor TF 215 with nominal torque 2000 Nm and possible 200% overload. The position is measured with Renishaw high accuracy optical rotary encoder based on stainless steel ring and two TONiC series incremental encoder readheads. It provides a basic resolution of 40 000 ppr and with possible interpolation, it is up to 8 000 000 ppr. Both sensors are coupled with the common shaft of the testing system. Temperature is measured on multiple locations of the actuator with PT 100 temperature sensors. Measured data are processed and displayed by the measuring station NI PXI with the graphical user interface which provides full control over the measurement process. The user interface is shown in Fig.3. The load cycle of the system is defined by the spreadsheet including time dependence of torque, speed, and position. This spreadsheet is created using MS Excel and is uploaded to the interface before measurement starts. Uploaded data are converted to the required form and subsequently sent to the PLC which provides the control of both machines. Defining the load cycle with a spreadsheet allows us to easily apply almost any type of load on the common shaft. The graphical interface also allows user to generate a report from past measurements.

V. Conclusion

This paper has described a laboratory system for investigating the performance of high precision actuators with a special approach of determining the load cycle by spreadsheet. The use of this type of data input makes the implementation of the load cycle simpler and allows user to easily determine the parameters of the cycle. The future work will be focused on creating an application to automatically prescribe a load cycle from real measured data and methodology proposal for simulation of the loads with a fast-changing moment of inertia. This type of load cycle implementation can be enhanced in the future for use in the more complex system - emulator of dynamic loads.

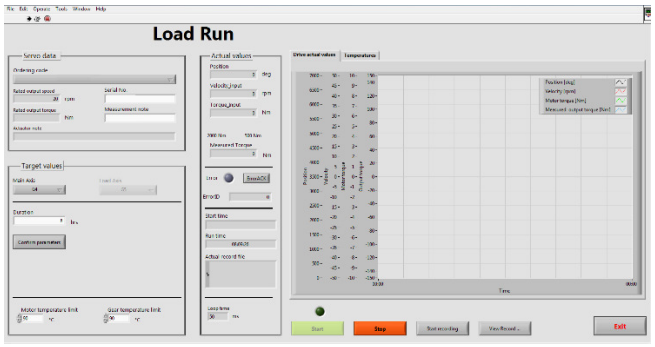


Fig.3 User interface

VI. Acknowledgment

This work was supported by the Slovak Research and Development Agency under contract No. APVV-15-0750 and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under the project VEGA 1/0493/19.

References

[1] E. J. Thomas and P.W. Sauer, “Engine-simulation control of an induction motor,” in *Proc. IEEE Elect. Insul. Conf. Elect. Manuf. Coil Winding Conf.*, 1999, pp. 589–594.

[2] C. Gan, R. Todd, and J. M. Apsley, “Drive system dynamics compensator for a mechanical system emulator,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 70–78, Jan. 2015.

[3] A. Sajadi, Ł. Rosłaniec, M. Kłos, P. Biczel, and K. A. Loparo, “An emulator for fixed pitch wind turbine studies,” *Renewable Energy*, vol. 87, pp. 391–402, Mar. 2016.

[4] P. Fajri, V. A. K. Prabhala, andM. Ferdowsi, “Emulating on-road operating conditions for electric-drive propulsion systems,” *IEEE Trans. Energy Convers.*, vol. 31, no. 1, pp. 1–11, Mar. 2016.

[5] R. J. Cavagnaro, J. C. Neely, F. Fay, J. L. Mendia, and J. A. Rea, “Evaluation of electromechanical systems dynamically emulating a candidate hydrokinetic turbine,” *IEEE T rans. Sustain. Energy*, vol. 7, no. 1, pp. 390 – 399, Jan. 2016

Geometrically Frustrated Magnetic Systems under High Pressure

[1]*Július BAČKAI (1st year)*
Supervisor: [2]Slavomír GABÁNI

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic
[1,2]Centre of Low Temperature Physics, Institute of Experimental Physics Slovak Academy of Sciences, Slovak Republic

[1]julius.backai@tuke.sk, [2]gabani@saske.sk

***Abstract*—Rare-earth tetraborides (REB_4) crystallize in tetragonal crystal structure, where the sublattice of rare-earth ions forms in *ab*-plane a geometrically frustrated Shastry-Sutherland lattice. During my postgradual studies I am going to construct detailed T_N vs. B phase diagram of three tetraborides (TmB_4, ErB_4 and HoB_4) from data of temperature and field dependencies of resistance, magnetization and heat capacity. We will apply hydrostatic pressure up to 15 GPa using a piston cylinder cell as well as diamond anvil cell in temperature range 1.8K – 300K and in magnetic field up to 6T to our samples. The obtained results will show us shift of ordering temperature T_N as well as shifts of boundaries between different magnetic phases. The effect of pressure on magnetic ions in those tetraborides will be discussed and compared between studied tetraborides.**

***Keywords*—geometrically frustrated magnetic systems, high pressure, tetraborides**

I. Introduction

Magnetic systems with geometrically frustrated interactions have attracted widespread interest in the last years due to discoveries of various new types of complex ground states [1]. The interest in such systems, in particular, is related also with magnetization plateaus at fractional values of saturation magnetization. It was shown that rare earth tetraborides (REB_4), with tetragonal crystal structure, represent 2D frustrated magnets on the Shastry-Sutherland lattice (SSL), which are relatively easy and fully accessible for experiments up to the saturation magnetic field.

Probably the most studied among tetraborides is TmB_4, which exhibits a rich phase diagram as a function of temperature and field with T_N = 11.7 K and J = 6 (total angular momentum of Tm^{3+}). [2], [3] During my postgraduate study I am studying TmB_4 together with another two tetraborides ErB_4 and HoB_4. The main purpose of my thesis is to explore influence of high pressure on magnetorezistivity, heat capacity and magnetization as well as on changes in the magnetic phase diagram of these compounds.

II. Geometrically Frustrated Systems

Research of geometrically frustrated magnetic systems has supported new theories dealing with interaction of magnetic spins, which indicate strong anisotropy. The same studies contribute to development of new materials in all classes of magnetic systems. Based on site disorder and frustration we can divide magnetic materials into four categories (Fig. 1). As we can see on Fig. 1 there are two categories which exhibit frustration: spin glasses and geometrically frustrated systems.

It is important to know the differences between those two kinds of materials. In the majority of spin glasses the reason of site disorder is connected to frustration. Very simple and clear example is illustrated in Fig. 2 top.

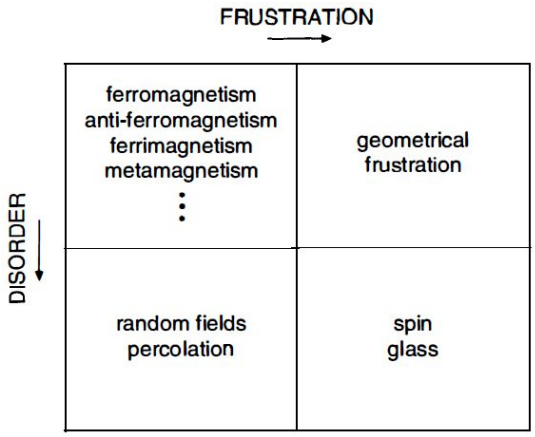


Fig. 1. Ground states of insulating magnets from the perspective of site disorder and frustration. [4]

If we modify lattice where are originally exist only antiferromagnetic bonds (AF) by random spreading spins with ferromagnetic bonds (FM), it creates site disorder which leads to frustration. Another type of frustration occurs if the lattice itself is responsible for frustration Fig. 2 bottom. Due to triangular symmetry we can observe quantitative same geometrically frustration as in spin glasses but without necessity of spin disorder. [4] Further we will be discussing only geometrical frustration systems which show low spin disorder but strong geometrical frustration.

Importance of study of geometrically frustrated systems has shown in supersolid phases research. This new type of matter has property of both solid crystal and superfluid substance. Whereas there is still no convincing experimental realization of supersolid matter, geometrically frustrated systems appear like right candidate for modeling this new sort of matter. [5]

Geometrically frustrated systems hold another important role in searching for skyrmion crystals. Those extraordinary objects, whose spins are oriented in all directions on the

sphere surface, attracted attention in last years due to their potential usage in future memory technology with remarkable properties such as nano size scale, high stability or ultralow critical drive current. Skyrmions were already discovered and theoretically confirmed in chiral magnets. It is assumed that next suitable systems for generating skyrmions close room temperature are geometrically frustrated systems with Shastry-Sutherland (S-S) lattice (Fig. 3). [6] Well known representatives of magnetic materials with S-S lattice are rare earth tetraborides, as already mentioned TmB_4, ErB_4 and HoB_4.

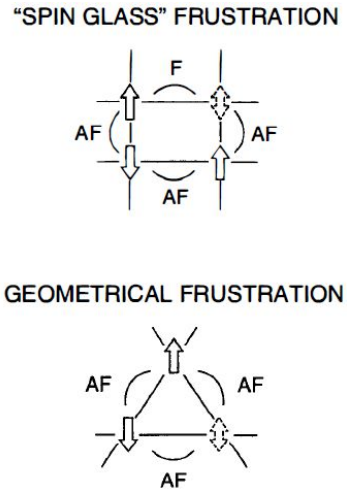


Fig. 2. Top plaquette of the square AF lattice with one bond replaced by a FM bond, illustrating frustration induced by site disorder common to most spin glasses. Bottom plaquette of the triangular lattice, showing frustration occurring among spins in a site-ordered system. [4]

III. TETRABORIDES

The rare earth tetraborides crystallize in the tetragonal structure P_4/mbm, where the boron network is formed by chains of octahedral along the c-axis. In the c-plane four octahedral chains are linked by a pair of boron atoms. While one additional electron is available in the conduction band, so the tetraborides are good metals.

Along the c-axis the rare earth ions are fixed in the *a*-*b*-plane (Fig. 3) between $B_6 – B_2$ layers. In the c-plane the ions are surrounded by boron octahedral. In doing so, the lanthanoid inos form a square in the *a*-*b*-plane, which is rotated by a small angle around the *c*-axis. Each ion has two nearest neighbors on the square and two in the neighboring square. One next nearest neighbor is located in the neighbor unit cell nearly in the same distance. Starting from this structure it is possible to map the rare earth ion grid to the Shastry-Sutherland lattice.

In 1981 Siriam Shastry and Bill Sutherland established a model which describes the behavior of antiferromagnetic Heisenberg spins in consideration of nearest and next nearest neighbor interactions in two and three dimensions. [7] The featured of the model is that exact results for the ground state could be found for a wide range of parameters.

The moments, located at nodes of square lattice, interact with their neighbors via the dashed diagonal bonds *J* and via black bonds *J'* with their next nearest neighbors (Fig. 3). *J* and *J'* are the coupling constants. The model shows that for a large interaction on the bond *J'* the systems is similar to a normal Heisenberg antiferromagnet with Néel ground state, but for a strong coupling on the *J* bond the system behaves in a different way. The lattice build dimmers on the diagonal bonds, which are antiferromagnetical coupled and the ground state of each dimer is a singlet. Shastry and Sutherland called this phase a "quantum spin liquid" because the interaction of the states is liquid like with only short range order. The excited triplets on a dimer can hop between the *J/J'* ratio and the spin *S*.

As the ground state of many-particle systems is usually not exactly known, the model was interesting from the theoretical point of view. Without any analogous material, the model was without any importance in experimental physics up to end of the nineties. In 1999 Kageyama et al. [8] studied the magnetic properties of the $SrCu_2(BO_3)_2$ which was the first example of the Sastry-Sutherland lattice.

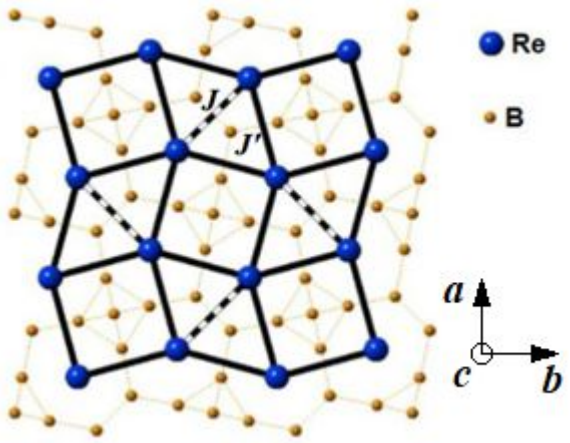


Fig. 3. The Shastry – Sutherland lattice. *J* and *J'* are the exchange energies on the bonds. [9]

As already mentioned, the rare earth tetraborides are good metals, therefore one expects the RKKY exchange interaction to be relevant between the localized moments.

The RKKY exchange is an indirect interaction, which is mediated by the conduction electrons. The localized magnetic moments polarize the surrounded conduction electrons. These polarized electrons interact with other localized magnetic moments. This leads to an indirect coupling between the localized moments. The exchange energy depends on the distance between the magnetic ions and the Fermi surface radius. The magnitude of the energy decreases with the distance between the ions and the sign alternates. Therefore the interaction can be ferromagnetic or antiferromagnetic. [10]

IV. EXPERIMENTAL METHODS

To acquire all necessary data for fulfillment every goal in my PhD thesis I will be using several experimental methods. Major experiments will be realized in helium-4 cryostat together with high pressure cell.

Helium 4 cryostat

The temperature range 1.3 K < *T* < 4.2 K is determined by the normal boiling point of ^{4}He and the temperature at which its vapor pressure has become very small. There are two ways to reach this temperature range using pumped vapors above liquid ^{4}He.

The simplest way is just pump the vapor above the liquid

^{4}He bath away to decrease its temperature. This is very uneconomical because about 40% of the liquid ^{4}He has to be evaporated to cool it from 4.2 K to 1.3 K due to large change of its specific heat in this temperature range. On the other hand, the specific heat of solids is rather small in this temperature range. To cool them from 4.2 K to 1.3 K we have to evaporate only a small fraction of liquid ^{4}He. It is therefore much more efficient to leave the main part of the liquid at its normal boiling point of 4.2 K and just pump on a small fraction of it in a separate container to reach the lower temperature for experiment. This type of cryostat is located in laboratory at Centre of Low Temperature Physics, Institute of Experimental Physics of the Slovak Academy of Sciences in Košice.

The design of continuously operating ^{4}He evaporation cryostat is on Fig.4. In such a refrigerator a small fraction of the liquid from the main 4.2 K bath flows through a suitable flow impedance into a small vessel of several cm^3. Through the central tube we pump on the liquid arriving in this evaporation vessel. The liquid from the main bath at 1 bar is enthalpically expanded trough the impedance and will arrive at a lower temperature in the evaporation vessel. Again, almost half of the heat of evaporation is used for cooling the liquid, the other half can be utilized to fill up the inner vessel with liquid and to cool sample. This vessel will continue to fill up the inner vessel with liquid and to cool something else. This vessel will continue to fill until the level of the liquid in the pumping tube is at height at which the heat transferred from the main helium bath through this column of liquid, plus the heat from the experiment, just balances the cooling power of the refrigerator available from the latent heat L of evaporation.

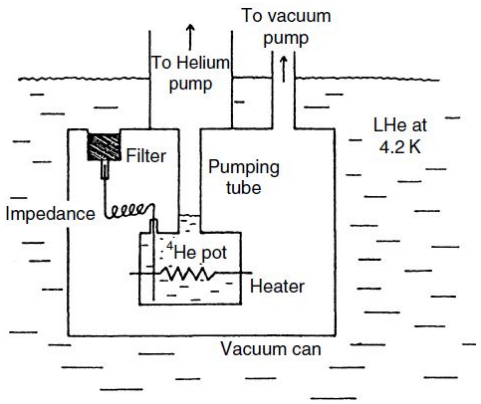


Fig. 4. A continuously operating ^{4}He refrigerator for the temperature range between 1.3 K and 4.2 K. [11]

The refrigerator is self-regulating, if we increase the external load, the level of the liquid in the pumping tube will drop, so its contribution to the heat transferred to the inner vessel is reduced. This experimental equipment create basis for our current and future experiments. [11]

Piston cylinder cell

In addition to the now common diamond anvil cells (DACs), hydrostatic piston cylinder cells have a large field of application where large volume and electrical input are necessary. The physics of semiconductors and superconductors, in particular, is a field of rather wide application of hydrostatic pressure. The electron energy spectrum and carrier concentration strongly change under pressure and can be investigated by transport and optical methods. Therefore for the majority of physical applications a hydrostatic cell has to have an electrical or an optical input or both and it has to be non-magnetic, and quite small.

Fig. 5. Piston cylinder cell from Centre of Low Temperature Physics, Slovak Academy of Sciences in Košice. Top pictures shows detail on sample, which maximum dimension is about 350μm on bottom picture is disassembled whole piston cylinder cell.

From the point of view of the pressure limit the cylinder is obviously not the weakest part of the cell, in spite of the fact that it works in tension. The cell also includes a piston, plugs, seals and electrical leads. The choice of materials for specific applications, such as low temperatures, magnetic fields or neutron diffraction is limited. The cell must be reliable and safe. It is desirable to have a simple cell which does not need much attention and is suitable for not very specialized experiments. Clamped cells of small dimensions as possible are preferable for easy combination with cryostats, solenoids and other set-ups. [12]

A particular actual cells has to fulfill these requirements. Therefore the working pressure limit of 3-3.5 GPa may be expected for materials and designs available now. Piston cylinder cell with those properties which is suitable for very low temperature experiments (Fig. 5) is situated at Centre of Low Temperature Physics, Institute of Experimental Physics of the Slovak Academy of Sciences in Košice.

V. Experimental results

Measurement of pressure inside cells is also very important for high pressure experiments. In DAC or in piston cells with transparent window it could be done by using small pieces of ruby. Then by shift in light spectrum we can find amount of pressure in cell. This method is contactless and very simple, but it works only if area of high pressure has a transparent properties.

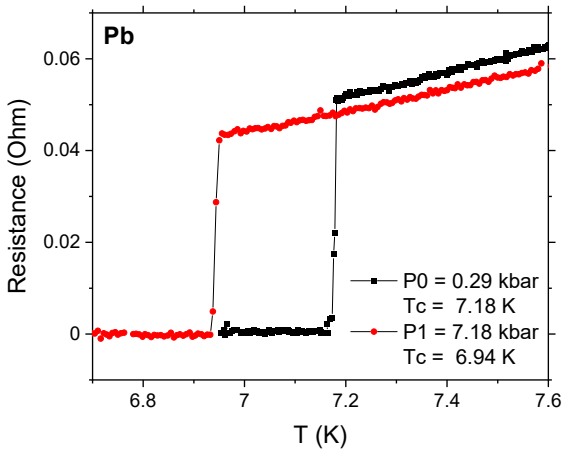


Fig. 6. Superconducting phase transition of lead at atmospheric pressure and under pressure of 7 kbar in piston cylinder cell.

In our pressure cell (Fig. 5) and others which do not have transparent window, we can use well known element and measure shift in superconducting phase transition temperature. In Fig. 6 is shown how pressure affects superconducting phase transition of lead. Through equation [13]:

$$p\ (\mathrm{GPa}) = \frac{7.19 - T_C}{0.348}$$

where p is pressure in piston cylinder cell and T_C is temperature of phase transition of lead under pressure, we can calculate the value of pressure locked in cell. We use this equation in pressure measurement in piston cylinder cell (Fig. 6) by very pure lead sample, and we successfully generate pressure p = 7.18 kbar. After test measurements we are ready to insert our tetraborides samples into piston cylinder cell and perform measurements.

VI. Conclusion

The experimental study of geometrically frustrated systems has a huge potential not only for basic research but also for applications. The future of information technologies is in quantum computers and geometrically frustrated tetraborides with S-S lattice seems to play important role in it. During my studies I will try to shed more light to the investigation of tetraborides and their phase diagrams which shows several different magnetic phases under extreme conditions such as very low temperatures, pressure and magnetic field.

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 contracts of Slovak agencies VEGA, grant no. 2/0032/16 and 2/0032/20, APVV, grant no. 17-0020 and DAAD project no. 57452699. Liquid nitrogen for the experiments has been sponsored by U.S. Steel Košice, s.r.o.

References

[1] J. Trinh, S. Mitra, Ch. Panagopoulos, T. Kong, P.C. Canfield, and A. P. Ramirez: Physical Review Lettters. 121, 167203 (2018)
[2] K. Siemensmeyer, E. Wulf, H.-J. Mikeska, K. Flachbart, S. Gabáni, S. Maťaš, P. Priputen, A. Efdokimova, and N. Shitsevalova: Physical Review Letters 101, 177201 (2008)
[3] Mat. Orendáč, S. Gabáni, E. Gažo, G. Pristáš, N. Shitsevalova, K. Siemensmeyer, and K. Flachbart: Scientific Reports, 8 10933 (2018)
[4] A. P. Ramirez: Annual Reiew of Materials Science, 453-80 (1994)
[5] K. Wierschem, Z. Zhang, A. Wibawa, and P. Sengupta: European Physics Journal B, 91, 201 (2018)
[6] Y. Jihane, L. Weihao, H. Zongpeng, L. Jinjing, Ch. Jun, Ch. Deyang, H. Zhipeng, and Q. Minghui: Physics Status Solidi RRL. 1900161 (2019)
[7] B. S. Shastry, and B. Sutherland: Physica B+C, 108(1 – 3), 1069 – 1070(1981).
[8] H. Kageyama, K. Yoshimura, R. Stern, N. V. Mushinkov, K. Onizuka, M. Kato, K. Kosuge, C. P. Slichter, T. Goto, and Y. Ueda: Physical Review Letters, 82, 3186 (1999).
[9] D. Brunt, G. Balakrishnan, D. A. Mayoh, M. R. Lees, D. Gorbunov, N. Qureshi, and O. A. Petrenko: Scientific Reports, 8, 232 (2018)
[10] J. J. Feng, L. Huo, W.C Huang, Y. Wang, M. H. Qin, J.-M. Liu, and Z. Ren: Europhysics Letters. 105, 17009 (2014)
[11] Frank Pobell. Matter and Methods at Low Temperatures (Third Edition). ISBN-10 3-540-46356-9
[12] Mikhail Eremets. High Pressure Experimental Methods. ISBN-0-19-856269-1
[13] R. I. Boughton, J. L. Olsen, and C. Palmy: Progress in Low Temperature Physics. 6, 163 (1970)

A general framework for scientific computing

[1]*Daniel GECÁŠEK (2nd year),*
Supervisor: [2]Ján GENČI, Consultant: [3] Pavol Bobík

[1,2]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic
[3]Institute of Experimental Physics, Slovak academy of sciences, Slovak Republic

[1]daniel.gecasek@tuke.sk, [2]jan.genci@tuke.sk, [3]bobik@saske.sk

***Abstract*—A general framework for scientific computing could improve the impact of ad-hoc developed scientific models. Thus research focused on design and system requirements for such framework could help to spread the popularity of not very popular but valuable scientific models. Thus the incremental design of such a system based on first case studies is a good way to attack the problem.**

***Keywords*—data analysis, scientific computing, optimization, visualization**

I. Introduction

This work uses a model for the evaluation of cosmic ray trajectories as a case study and the first application for the wider framework to be developed to automate scientific models. The cosmic rays were discovered more than a century ago [1] in 1912. The early age of trajectory simulations starts with manual solutions, but very soon computers are used. In [2] authors used "Bush's differential analyzer" [3] to investigate allowed cones of cosmic radiation.

Computer power increased continuously in the next decades and tracing techniques become widely used to explain cosmic ray observations in the magnetosphere. However, at the same time, the model of the geomagnetic field becomes more complex and complicated for numerical evaluation. Thus the whole task becomes more and more complex and numerically expensive. The first authors, that used a model of the geomagnetic field for trajectory simulations were Gall and coauthors [4] in 1968 using Williams and Mead model of the geomagnetic field.

The task is still computationally expensive. A huge report from 2020 [5] with rigidity tables covering the whole globe in last decades state "The extreme requirement of intensive computation to obtain a sufficient number of particle trajectories to evaluate cosmic ray access to a specific location on the earth or in the earth's magnetosphere may involve obtaining solutions to millions of individual cosmic ray trajectories. Therefore efficient computation is essential (and a very fast computer desirable)."

II. State of the art

The COR system, for evaluation of cosmic ray trajectories in the geomagnetic field, was developed in cooperation with the Technical University of Košice (TUKE) and Institute of Experimental Physics Slovak Academy of Sciences (IEP SAS) in the last 4 years. The system typically traces one million trajectories for evaluation of cosmic ray intensity at the requested location, which takes 10-20 hours on a node with 32 cores. The COR model as an example of a computationally expensive model is a suitable case study for the design specification of a framework that aims to simplify the automation of scientific models.

III. COR system for cosmic ray trajectories in Earth magnetosphere simulation

We presented an article about the COR system at the 20th conference of Czech and Slovak physicists [6]. In the article, we mainly described a new module for historical effective cutoff rigidity evaluated in period 1-1901 CE. We also described the following features of the system:

- Our efforts in optimizing the simulation runtime by creating formulas to estimate starting rigidity just below lower cutoff rigidity.
- Simulations of the multidirectional cosmic ray trajectories.
- Visualization of cosmic ray particle trajectory. In the figure 1 we can see an example of a 3D view selected from the available projections and 3D view.

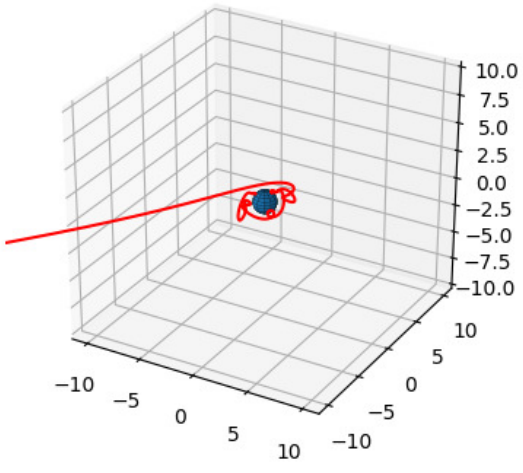


Fig. 1. Visualization of cosmic ray trajectory in GSM coordinates in Earth radii (Re) units.

The main added feature was a module for historical effective cutoff rigidity evaluation. The module allows users to evaluate magnetosphere transparency (cut off rigidity and spectrum of allowed and forbidden rigidities) for cosmic rays in the last two millennia. In the current version of the module, trajectories are simulated only for vertically incoming particles.

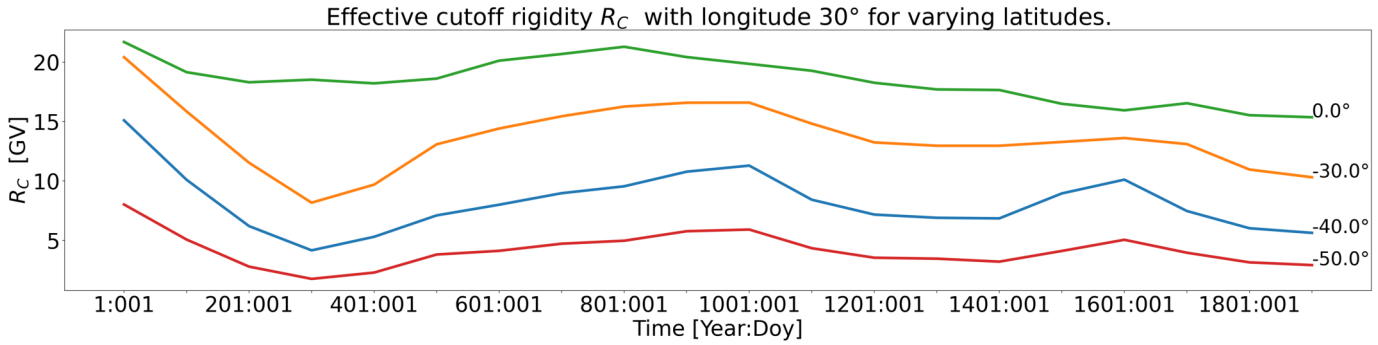


Fig. 2. Visualization of effective cutoff rigidities for years 1 to 1901 with a 100 year interval.

In figure 2 long term changes of effective vertical cutoff rigidities at the longitude 30° for selected latitudes of the southern hemisphere for years between 1 and 1901 CE are presented. The time of calculation is 1st January 00:00 for each point in time. The inspection of the figure shows that Earth's magnetosphere is more transparent for cosmic rays now than it was two thousand years ago. Cutoff rigidities were higher in first century than now. The figure presents only selected positions, but the trend is general for the whole Earth surface. Further, inspection shows that some points could see opposite trends for some period of times, and thus could be relevant to investigate local changes of magnetosphere transparency for cosmic rays and consequences to carbon 14 (^{14}C) production and distribution.

Another notable activity was the optimization of starting rigidity. In the multidirectional simulation, all directional simulations in COR 1.0 started at 0.01 gigavolt (GV). With this optimization, the following formulas to estimate starting rigidity for any direction of any point were derived:

- Formula based on vertical cutoff rigidities, used only for simulations of vertical incoming directions.
- Formula based on simulations along the selected meridian.
- Formula based on simulations covering whole Earth's surface with selected steps in latitude and longitude.

As a result, the optimization of running times described in Table I was achieved.

TABLE I
TABLE OF DURATIONS OF SIMULATION FOR SELECTED POSITIONS WITH OPTIMIZATIONS TURNED ON OR OFF.

Latitude	Longitude	Optimization	Duration
−60.15°	287.79°	No	14:04:31
−60.15°	287.79°	Yes	09:12:33
−10.05°	287.79°	No	05:05:36
−10.05°	287.79°	Yes	04:24:15
40.21°	287.79°	No	14:11:44
40.21°	287.79°	Yes	10:28:57

IV. Future Research

In the future we plan to examine suitability of existing standard data formats that could be used in our proposed system. Candidate file formats that we plan to examine further are CSV [7], JSON [8], XML [9], CDF [10], NetCDF [11], HDF [12] and FITS [13].

We also plan to examine magnetic field models based on their runtime performance and compare their result values to each other. The models we plan to compare are:

- Newest International Geomagnetic Reference Field (IGRF) implemented in Python
- Older version of IGRF implemented in Fortran-77.
- CHAOS-7 implemented in Python
- Tsyganenko models of external geomagentic field implemented in Fortran-77 (T89 [14], T96 [15] and TS05 [16]).

The test of the system should be proven in the simulation of historical cutoff rigidities for a grid of points with a resolution of 1° in both latitudinal and longitudinal directions for the last two thousand years. The meaning of this simulation is to find the distribution of ^{14}C isotope produced by cosmic rays in the atmosphere for different historical epochs, and consequently, to elaborate necessity of modification of current ^{14}C dating models [17].

References

[1] V. F. Hess, "Uber beobachtungen der durchdringenden strahlung bei sieben freiballonfahrten," *Phys. Zeits.*, vol. 13, pp. 1084–1091, 1912.
[2] G. Lemaitre and M. Vallarta, "On the geomagnetic analysis of cosmic radiation," *Physical Review*, vol. 49, no. 10, p. 719, 1936.
[3] V. Bush, "The differential analyzer. a new machine for solving differential equations," *Journal of the Franklin Institute*, vol. 212, no. 4, pp. 447–488, 1931.
[4] R. Gall, J. JiméNez, and L. Camacho, "Arrival of low-energy cosmic rays via the magnetospheric tail," *Journal of Geophysical Research*, vol. 73, no. 5, pp. 1593–1605, Mar. 1968.
[5] K. Copeland, *CARI-7 Documentation: Geomagnetic Cutoff Rigidity Calculations and Tables for 1965-2010*, 2020.
[6] D. Gecášek, J. Genci, and P. Bobík, "Cor system for cosmic rays trajectories in magnetosphere simulation," in *20th Conference of Czech and Slovak Physicists, Proceedings*. Slovak Physical Society, Czech Physical Society, 2020, pp. 163–164.
[7] Y. Shafranovich, *Common format and MIME type for comma-separated values (CSV) files*, SolidMatrix Technologies, Inc., 10 2005.
[8] T. Bray *et al.*, *The javascript object notation (json) data interchange format*, 3 2014.
[9] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau *et al.*, *Extensible markup language (XML) 1.0*, 10 2000.
[10] *CDF User's Guide*, 11 2020.
[11] R. Rew and G. Davis, "Netcdf: an interface for scientific data access," *IEEE computer graphics and applications*, vol. 10, no. 4, pp. 76–82, 1990.
[12] *HDF5 User's Guide*, 6 2019.
[13] D. C. Wells and E. W. Greisen, "Fits-a flexible image transport system," in *Image Processing in Astronomy*, 1979, p. 445.
[14] N. A. Tsyganenko, "A magnetospheric magnetic field model with a warped tail current sheet," *Planetary and Space Science*, vol. 37, no. 1, pp. 5–20, 1989.
[15] ——, "Effects of the solar wind conditions on the global magnetospheric configuration as deduced from data-based field models," in *International conference on substorms*, vol. 389. ESA Publication Division, 1996, pp. 181–185.
[16] N. A. Tsyganenko and M. I. Sitnov, "Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms," *Journal of Geophysical Research: Space Physics*, vol. 110, no. A3, 2005.
[17] W. Libbey, "Radiocarbon dating," *University of Chicago Press*, 1955.

Convolutional neural networks for chest x-ray medical imaging

[1]*Matej GAZDA (2nd year),*
Supervisor: [2]*Ján Plavka*

[1,2]Dept. of mathematics and theoretical informatics, FEI TU of Košice, Slovak Republic

[1]matej.gazda@tuke.sk, [2]jan.plavka@tuke.sk

***Abstract*—This paper provides overview of medical imaging and summarizes available datasets for the description of the chest x-ray image. It also includes recent work in the field with state-of-the-art methods and result of self-supervised learning.**

***Keywords*—biomedical engineering, chest-xray, convolutional neural networks, medical imaging**

I. Introduction

Medical imaging is considered a key contribution to the patient's well-being and overall health-care. Medical imaging utilization increases rapidly, up to 50% more than the last decade [1]. Computer tomography (CT), X-ray, sonography, and magnetic resonance imaging are all considered medical imaging techniques. We divide the tasks solved on medical imaging into localization, detection, classification, and segmentation. Localization is a process of finding a square bounded object on the image. Detection is a process similar to localization; however, we find the bounding for multiple objects, where each object can be a different class. Segmentation is another popular task, where the process aims to finds and highlights the contours of desired objects. Finally, classification is the most popular process to correctly evaluate the patient's current health state (healthy/ill).

Several studies focused on automatic chest X-ray description for diagnosis of viral or bacterial pneumonia induced by COVID-19, Tuberculosis, etc. Apostolopoulos et al. [2] classified COVID-19 off chest x-ray images using models pretrained on ImageNet. Similarly, Minaee et al. [3] used Resnet18, ResNet50, Squeezenet, and Densenet models pretrained on ImageNet to solve the COVID-19 classification task. So far, the published studies focused on the classification of respiratory diseases from chest x-ray utilized the deep neural network models either trained from scratch or pretrained from natural image datasets such as ImageNet. We focus on self-supervised pretraining on pretext task solved on big unlabeled chest x-ray dataset and transferring the knowledge to downstream task.

II. Overview of chest X-ray datasets

Classification of chest x-ray by convolutional neural networks (CNN) is a very studied and promising task. Several chest x-ray databases were created and labeled composed of tens of thousands of images, including CheXpert [4], Chest-xray14 [5], CELL [6], and more specified on COVID-19 COVIDGR and C19-Cohen [7]. We also include COVIDX database [8] created as a combination of the C19-Cohen and CELL database.

A. CheXpert

CheXpert [4] is a collection of 224316 chest x-ray images of 65240 patients. Patients are labeled for the presence of 14 most common chest x-ray observations as No finding (healthy), Support devices, Fracture, Cardiomegaly, Pleural Other, Pleural Effusion, Pneumothorax, Edema, Consolidation, Pneumonia, Lesion, Atelectasis, enlarged cardiom, and lung opacity. Labels were extracted from the free-text radiology reports by natural language processing tools.

B. Chest-xray 14

Chest-xray14 [5] is another dataset of chest x-ray images composed of 108948 frontal-view X-ray images from 32717 patients, and each patient is labeled by one or multiple pathology keywords. Again, similarly to CheXpert, the labels were automatically extracted from radiologists' reports by natural language processing tools.

C. CELL

Another dataset consisting of 5232 chest x-ray images from children was published in the prestigious journal CELL [6]. Out of 5232 chest x-ray images, 3883 was depicted as pneumonia (1345 viral and 2538 bacterial) and 1348 healthy patient. Note the chest x-ray images are only from children, and those images differ in comparison with Chest-xray 14 or CheXpert databases.

D. C19-COHEN

C19-Cohen [7] is one of the first publicly available datasets of chest x-ray images labeled as covid-19 and other pneumonia types. COVID-19 can be diagnosed based on x-ray or CT images of the chest. C19-Cohen consists of 481 samples where 320 patients were suffering from COVID19, and others suffered from other kinds of illness such as Legionella, Influenza, or bacterial pneumonia.

E. COVIDGR

COVIDGR [9] is the second dataset of chest x-ray images of healthy patients and patients suffering from the COVID-19. COVIDGR dataset contains 426 chest x-ray images of COVID-19 patients with four different severity levels and the same number of control subjects.

TABLE I
PERFORMANCE OF DEEP LEARNING MODELS ON DIFFERENT CHEST X-RAY DATASETS

Reference	Dataset	Task	Accuracy
Pham et al. [14]	CheXpert	Cardiomegaly	91%
Yao et al. [15]	Chest-xray14	Cardiomegaly	85.8 %
Kermany et al. [6]	CELL	Pneumonia	96.8%
Wang et al. [8]	COVIDX	COVID-19	99.7%
Tabik et al. [9]	COVIDGR	COVID-19	76.16 %
Our results	CELL	Pneumonia	95%

III. PROPOSED METHODS

CNNs are achieving state-of-the-art results for computer vision tasks for both natural and medical images. On the other hand, they are data-hungry, and therefore, their use on medical images is a bit harder as the image needs to be annotated by experts in the domain. On top of that, they are still not extensively used in production, as various concerns are raised by studies like [10] and [11]. Oakden [11] claims that even larger databases contain false labels that were extracted automatically from radiology reports. The solution to that would be to adhere to the TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) [12] statement, which introduces requirements for applying artificial intelligence models in clinical practice.

We explore the possibility of self-supervised pretraining in the medical domain. The representations are first learned in the self-supervised fashion on the pretext task and further reused on the the downstream task. The quality of features are evaluated by t-SNE algorithm and by the model's accuracy achieved on the various target datasets. Our work is summed up in [13], where we propose self-supervised neural network.

Table I depicts results of various tasks for 4 datasets: CheXpert, Chest-xray 14, COVIDGR and COVIDX. For CheXpert and Chest-xray 14, we chose to report the task of diagnosis of cardiomegaly. To evaluate the quality of the model's prediction, accuracy metric was used and it can be defined as the ratio of correctly predicted observations to number of all the observations.

It seems that the results provided by [14] and [15] are a bit different, but the distinction is not so huge. We hypothesize that this may be caused by the variety of the testing parts of both datasets and slightly different deep learning approaches.

The model provided by [6] correctly diagnose pneumonia on CELL dataset of chest x-ray images with accuracy over 92%. Surprisingly, the achieved precision is far better than other studies detecting pneumonia on datasets like CheXpert and chest-xray 14, where they achieved 65.8% ACC [5], but in multi-class settings. The age limitation of the patients in CELL dataset may cause this difference. Our self-supervised neural network achieved with linear evaluation protocol on CELL dataset similar results to its supervised counterpart (95%), however with significantly lower number of labeled samples.

The results of the correct diagnosis of COVID-19 on COVIDX can be overly optimistic, as the captured chest x-ray images were taken in different settings, which may contribute to the final accuracy. Furthermore, Tabik et al. [9] created a model with 76.16% AUC metric on COVIDGR dataset.

IV. CONCLUSION AND FUTURE WORK

CNNs are achieving state-of-the-art results for computer vision tasks. One major drawback of the CNNs is the unavailability of huge labeled datasets. Recent self-supervised learning approaches provide quality of learned features on par with supervised counterparts. The idea is to implement a self-supervised model aimed at chest x-ray and use it either as a fixed feature extractor or fine-tune it over the specified dataset with some labels. Self-supervised pretraining could decrease the required number of labeled images significantly. We showed that self-supervised learning can also be used in the medical imaging domain but needs further research. Future work includes the more sophisticated self-supervised methods to capture the more important features out of the image. The neural network is considered a black box model; however, some visualization techniques like deep taylor decomposition or layer-wise propagation to explain its decision exist.

REFERENCES

[1] R. Smith-Bindman, D. L. Miglioretti, E. Johnson, C. Lee, H. S. Feigelson, M. Flynn, R. T. Greenlee, R. L. Kruger, M. C. Hornbrook, D. Roblin, L. I. Solberg, N. Vanneman, S. Weinmann, and A. E. Williams, "Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 1996-2010," *JAMA*, vol. 307, no. 22, pp. 2400–2409, 06 2012. [Online]. Available: https://doi.org/10.1001/jama.2012.5960

[2] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.

[3] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *Medical image analysis*, vol. 65, p. 101794, 2020.

[4] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 590–597.

[5] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[6] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[7] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[8] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[9] S. Tabik, A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Area, D. Charte, E. Guirado, J. L. Suárez, J. Luengo, M. A. Valero-González, P. García-Villanova, E. Olmedo-Sánchez, and F. Herrera, "Covidgr dataset and covid-sdnet methodology for predicting covid-19 based on chest x-ray images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3595–3605, 2020.

[10] G. Maguolo and L. Nanni, "A critic evaluation of methods for covid-19 automatic detection from x-ray images," *arXiv preprint arXiv:2004.12823*, 2020.

[11] L. Oakden-Rayner, "Exploring large-scale public medical image datasets," *Academic radiology*, vol. 27, no. 1, pp. 106–112, 2020.

[12] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod) the tripod statement," *Circulation*, vol. 131, no. 2, pp. 211–219, 2015.

[13] M. Gazda, J. Gazda, J. Plavka, and P. Drotar, "Self-supervised deep convolutional neural network for chest x-ray classification," 2021.

[14] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, "Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels," *arXiv preprint arXiv:1911.06475*, 2019.

[15] L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, "Weakly supervised medical diagnosis and localization from multiple resolutions," *arXiv preprint arXiv:1803.07703*, 2018.

Digitization and transformation of the electricity market

[1]*Jozef HUMENÍK (2nd year)*
Supervisor: [2]Jaroslav Džmura

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

[1]jozef.humenik@student.tuke.sk, [2]jaroslav.dzmura@tuke.sk

***Abstract*— This article deals the process of digitization and transformation of power engineering and the electricity market, which represents just a small part of the whole issue. He talks about the current market situation, elements of smart grids and analysis of technologies. At the end is a summary of current requirements and needs of participants on the electricity market.**

***Keywords*—Digitization, Electricity market, Smart grids, Smart technologies**

I. INTRODUCTION

The process of digitization and transformation of energy is a picture of the requirements of market participants, which consists of legal entities conducting business activities in this sector, state institutions and other persons. The whole process consists of a few positives, and of course a lot of hard work, which will achieve the expected goal. The digitization and transformation of energy requires a few amendments to be tabled in the areas of law at international and national level, in the financial sector, in the technical and IT fields (Cyber and Information security), but also in the field of psychology. In which it is necessary to focus on changing the thinking of market participants and preparing them for changes that are irreversibly approaching. Thus, there is a wide-ranging development of many areas of industry that are interconnected.

Part of the energy transformation is the digitization and transformation of smart grids. Smart Grids are basically automated control systems that monitor the parameters of the transmission and distribution network also with the help of elements of artificial intelligence. An important element of smart grids is smart metering systems (SMS), also known as "Smart Metering". The data collected by SMS have huge potential and wide use in several areas of industry, such as electricity, gas and water. By analysing the obtained data, it is possible to achieve efficient operation and management of the electricity system, optimization of costs of individual market participants or automation of the electricity system. With the growing volume of data, it is necessary to focus our attention on the area of cyber and information security, which are one of the crucial components of the security environment. Cyber security cannot be tied to an area of society, as its specificity has an impact on the functioning of society.

II. CURRENT MARKET SITUATION

The current structure of energy is slowly but surely changing. The energy sector is being transformed and the concept of "4-D" is becoming more and more important. You can hear the voices that this is the future of energy. This term conceals the following four "D's":
1. Decentralization of electricity generation;
2. Decarbonization - reducing the carbon footprint;
3. Digitization;
4. Deregulation. [1]

The heading of decentralization of electricity generation can be understood as DER - Distributed Energy Resources. These resources include:
• Wind, photovoltaic/solar and hydro power plants;
• Accumulators (batteries);
• Diesel generators;
• Cogeneration units;
• Charging infrastructure for Electromobility;
• Heat storage (underground and above ground);
• Fuel cells. [2]

Simply, all elements that are a source of energy can be considered distributed energy sources. The most widespread sources include photovoltaic power plants due to simple installation, low maintenance and relatively low acquisition costs for the solar panels and batteries themselves. Distributed energy sources can economically manage the demand for electricity. The new requirements of the energy market will require the combination of more burdens with electric cars, energy sources and batteries, with the possibility of adapting to demand. In California, for example, there was a regulation that all newly built buildings must contain photovoltaic panels, which will reduce power consumption from the grid, reduce transmission losses and at the same time the point of consumption in question will become the called active customer. As far as Europe is concerned, there are tendencies and requirements to increase the self-sufficiency of new buildings. High self-sufficiency of buildings can be achieved by using my own energy sources. Here it is possible to talk about the so-called energy communities and Microgrids. Based on European Union data, it is estimated that by 2050, almost 50% of consumption points will produce electricity from RES or DER. At the same time, 1/3 of the consumption points will be a member of some energy community. The principle will be to increase self-sufficiency and reduce

spending. Such energy communities can then merge into larger units called microgrids. In certain cases, such communities may subsequently operate for some time in island traffic, independent of the outside world. A business model of this type is often referred to as VPP - Virtual Power Plant. The model takes to a higher level, making it more advanced and gaining in popularity thanks to advanced technologies for load size prediction and aggregation. [3]

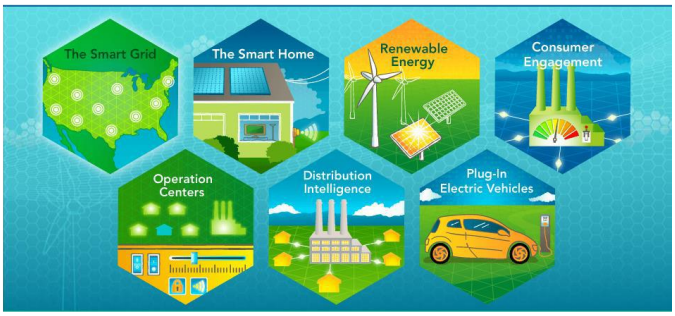


Fig. 1 An overview of key areas of electrical power engineering transformation [4]

III. Smart Technologies

The concept of smart grid creation is often composed of several technologies, which can be divided as follows:

- Smart Home;
- Energy management system;
- Microgrids;
- Automatic voltage control;
- Renewable Energy Sources (RES) and Electromobility;
- Intelligent transmission and distribution of electricity;
- Accumulation of electricity and distributed energy sources;
- Smart power stations or distribution transformer stations;
- Smart Measurement Systems (SMS) and Advanced Measurement Infrastructure (AMI). [5]

At the current stage of digitization and transformation of the energy sector as an industrial sector, the requirements and needs of market participants resonate, especially in a professional society, which can be considered as current challenges in this area. The current situation requires changes, especially in the area of:

a) telecommunications and IT solutions;
b) data exchanges between market participants;
c) creation of dynamic tariffs;
d) cyber and information security of critical and operational systems;
e) renewable and distributed energy sources;
f) building a charging infrastructure for Electromobility;
g) batteries;
h) the need to create new services;
i) legal norms.

The current situation, for example in the area of data exchange between market participants, is complicated. In order to simplify processes and save money, this situation needs to be resolved. It is necessary to devise a modern and smart way of exchanging data in order to achieve a system of data collection, sharing, processing and exchange. We currently know of 10 different parties who have access to energy data, whether measured or commercial. The modernization and transformation of energy has required the emergence of new services and market players, such as aggregation, flexibility, electricity storage or the construction of charging infrastructure. New market entrants include:

- Active customers, which represent a huge potential and a significant part of the market;
- Energy communities/Microgrids operators;
- Charging infrastructure operators;
- Operators of battery storage of electricity;
- Local resources;
- Aggregators (suppliers).

IV. Conclusion

Smart grids are the result of the digitization and transformation of the energy industry, which is an evolutionary process of modernizing electricity grids and which reflects the new requirements and needs of energy commodity market participants. The transformation of the energy industry consists of new requirements and needs for the production, transmission, distribution and operation of the electricity network, for the development of communication and IT technologies, Information and Cyber security and the harmonization of legal standards based on these inputs.

The digitization and transformation of energy should increase the security, quality and reliability of electricity transmission and distribution, information and cyber security, financial savings, increase the living standards of the population, increase the share of electricity production from renewable energy sources, greater environmental protection, e.g. reducing greenhouse gas emissions and using new technologies such as smart metering, electromobility, the smart home and so on.

At the top of the list of countries with the highest number of electric vehicles is China, which in 2018 had a registered 2.3 million electric vehicles and USA had 1.1 million cars. The number of vehicles in Europe was over 1.35 million, of which 1,570 passenger electric vehicles were registered in Slovakia (619 PHEV and 959 BEV). In addition to passenger cars, 74 light commercial vehicles and 47 electric buses were registered.

V. References

[1] Ministerstvo hospodárstva Slovenskej republiky: *Integrovaný národný energetický a klimatický plán na roky 2021 – 2030*. 2019. [Online]. Dostupné na: <https://www.mhsr.sk/uploads/files/zsrwR58V.pdf>

[2] IESO: *A Progress Report on Contracted Electricity Supply*. 2020. [Online]. Dostupné na: <https://www.ieso.ca/-/media/Files/IESO/Document-Library/contracted-electricity-supply/Progress-Report-Contracted-Supply-Q3-2020.ashx>

[3] LEHMBRUCK, L.; KRETZ, J.; AENGENVOORT, J.; SIOSHANSI, F.: *Chapter 10 - Aggregation of front- and behind-the-meter: the evolving VPP business model*. Editor(s): Fereidoon Sioshansi, Behind and Beyond the Meter, Academic Press. 2020. pages 211-232, ISBN 9780128199510.

[4] AJAZ, W.; BERNELL, D.: Microgrids and the transition toward decentralized energy systems in the United States: A Multi-Level Perspective. Energy Policy. Volume 149. 2021. 112094. ISSN 0301-4215.

[5] Ministerstvo energetiky Spojených štátov amerických. 2020. [Online]. Dostupné na: <https://www.smartgrid.gov/the_smart_grid/#smart_grid>

Explainability of artificial intelligence in medical domain

[1]*Michal KOLÁRIK (1st year),*
Supervisor: [2]*Ján PARALIČ, Consultant:* [3]*Martin SARNOVSKÝ*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEEI TU of Košice, Slovak Republic

[1]michal.kolarik@tuke.sk, [2]jan.paralic@tuke.sk, [3]martin.sarnovsky@tuke.sk

***Abstract*—This paper describes the current state in the literature in the context of the explainability of artificial intelligence with a focus on the medical domain. The paper provides a definition of interpretability and explainability and also the need for transparency and explainability in the medical domain. This work also provides desiderata for XAI methods. Methods of explainability are divided into four categories according to different approaches and briefly explained. Current open research problems in this area are identified.**

***Keywords*—explainability, interpretability, transparent AI, XAI**

I. Introduction

Recent Artificial Intelligence (AI) systems based on machine learning algorithms excel in many fields. AI can outperform humans in visual tasks or strategic games but also they become an indispensable part of everyday lives, such as online services that analyze our shopping carts or systems that allow us to make decisions based on data.

AI systems based on black-box models are used in many areas today. These systems used in smartphone applications or online services do not have key requirements for model transparency but focus mainly on model accuracy and cost. If such a model fails and, e.g. does not recognize the person logging into the system or the translation system makes a grammatical error in translation, it usually does not have major consequences. The requirements for transparency and trust in these applications are low. However, in applications that are critical to human safety, these requirements play an important role and can even be a decisive factor in deciding when deployment such a system. In particular, if the consequences of an AI decision can be life-threatening e.g. in autonomous cars or in the medical domain. Therefore, transparency is more important especially in these areas and, in addition to transparency, it is also necessary to build trust in the AI-based system that is deployed. In order for the model to be transparent, we need to know how it works and on the basis on which evidence it makes decisions. XAI methods provide tools that can help with these problems. Besides, there are legislative requirements for clarity and transparency in the processing of personal data as well as medical data.

II. Current state of XAI in Healthcare

In the healthcare field, artificial intelligence can be very beneficial. There are already practical deployments of AI to help doctors identify heart failure problems [1]. However, the full potential of AI systems is limited by the inability of algorithms to explain their results and decisions to human experts. This is a huge problem, especially in the medical field where doctors need to understand why AI has made a decision and how it came to that decision. In some cases, human life may depend on this decision. Transparent algorithms could reasonably increase the confidence of medical experts in future AI systems [2]. Therefore, research aimed at creating explainable AI (XAI) systems for medical applications requires maintaining a high level of learning performance for various machine learning techniques and human-computer interaction. There is a certain tension between the accuracy and explainability of machine learning methods. The most powerful models (especially deep learning or ensembles) are often least transparent and methods that provide clear and comprehensible explanations known as interpretable models (e.g. decision trees) are less accurate [3].

In the field of healthcare, the motivation for the XAI methods is clear. In many cases, both end-users and the critical nature of the predictions require some transparency, either for user involvement or for patient safety. However, sometimes an explanation of machine learning predictions is not enough. It is important to think about how the results are interpreted by the end-user, how they are incorporated into the work process, or how they are used in other ways. Healthcare experts are often overwhelmed by the influx of patients, the influx of data about these patients, and the related tasks that are required of them such as entering data into the system, electronic records, providing care. Therefore, if AI systems and their explainability are not presented in the right way, it will not help healthcare experts, but on contrary, it will be another task for them. Therefore, these systems should be created in terms of the domain and the perspective of the user who will work with them [2].

AI is often associated with the idea that artificial intelligence should replace the decisions of health professionals. However, it is not necessary to create systems in this way. Conversely, AI can be very helpful in important decisions that doctors must make, especially if the reasons for AI decisions or predictions are properly explained.

A. Explainability and Interpretability

Interpretability and explainability are often used in the literature as synonyms, but some works distinguish them. The term understanding is sometimes used as a synonym for interpretation and explanation in the context of the XAI [4]. In this

context, the term "understanding" usually means a functional understanding of the model, as opposed to an algorithmic understanding of the model at a low level. Understanding tries to characterize the behavior of a black-box model without trying to clarify its internal behavior.

In [5] authors distinguish between interpretation, which they define as the mapping of an abstract concept to a domain that can be perceived and understood by a human expert, and explanation, which they define as a set of interpretable area features that contributed to this example of decision making. In [6] Edwards and Veale divided the explanations into model-centric and object-centric terms that correspond to the definitions of interpretability and explainability from [5]. Similar tasks are explained in [7] as global and local interpretability.

The EU legislation and its General Data Protection Regulations (GDPR), which deals with the processing of personal data, mention only term of explainability. Comprehensibility [8] is used in the literature as a synonym for interpretability. Transparency [9] is used as a synonym for the interpretability of the model, which is in a sense an understanding of the logic of how the model works.

B. Explainability as part of next-generation AI systems

The concept of explainability is increasingly occurring along with the requirements for AI systems as one of the main requirements. These requirements may be part of the requirements for the application domain, such as banking, healthcare, or they may be part of legislative regulations that are gradually coming along with the development of AI systems. The ethical aspect should be equally important. They deal with the direct but also indirect impact of AI decisions on people's lives.

J. Fjeld and A. Naggy [10] in their study compared 36 important documents from various fields, such as organizations or government documents or recommendations for AI, and based on these documents defined 8 key principles of contemporary AI, including the terms explainability and transparency in the same theme:

- *Privacy*. AI systems should respect the right to privacy of individuals, both in the use of data in technology systems and in the provision of data to decision-making agencies.
- *Accountability*. It is important that responsibility for the impacts of AI systems is properly defined and that remedial action is provided.
- *Safety and security*. AI systems must be secure and operate as designed. They also need to be secured and resilient against abuse by unauthorized parties.
- *Transparency and explainability*. AI systems should be designed and implemented to allow supervision as well as interpretation of activities in comprehensible output and to provide information on where, how, and when these systems are used. This principle is the response to challenges such as transparency, explainability, open source data and algorithms, or right to information.
- *Fairness and non-discrimination*. The principles of justice and non-discrimination require that AI systems should be designed and used to maximize fairness and minimize bias.
- *Human control of technology*. This principle requires that important decisions remain under human control all the time.
- *Professional responsibility*. This principle addresses the responsibilities and role of individuals in the process of developing and deploying AI systems and calls for professionalism and integrity in ensuring communication with stakeholders on the long-term effects of these systems.
- *Promotion of human values*. The principles of human values state that the goals pursued by AI and how they are pursued should correspond with our values and generally support the human well-being.

In addition to these key principles, which should become part of modern AI systems, many scientists, lawyers and psychologists are currently dealing with ethical issues related to AI. Especially because with the increasing possibilities that AI offers us, new problems or questions arise, especially in applications that have a major impact on human lives. For example, how do we ensure that AI is fair and free from racial or gender prejudice? Who will be responsible if life is threatened as a result of AI's decision? How to ensure that AI is fair and transparent? When can the AI itself decide and when is it necessary to keep the supervision of the responsible person?

Recent initiatives in this area have also confirmed the importance of these problems. In the European Union, the AI Expert Group has produced document the Ethics Guidelines for Trustworthy AI [11], which provides guidelines for the development of trusted AI based on the principles of fundamental human rights that apply throughout the EU. The result is a kind of framework that defines 4 ethical principles:

1) *Respect for human autonomy* - A person has right to supervise the system and to intervene in the AI process at any time.
2) *Prevention of harm* - This principle aims to prevent AI systems from harming a person, whether in physical or mental form.
3) *Fairness* - The aim is to prevent discrimination or bias in AI.
4) *Explainability* - AI systems and their decisions should be explained in a way that is understandable to the stakeholders involved. Humans should know when they use the AI system and must be informed about its capabilities and limitations.

Also, commercial companies engaged in research in AI applications are interested in creating systems that should be ethical, fair, or transparent. For example, Google has released a document with its own principles that they want to follow when creating AI systems [12].

In the healthcare field, research focuses on the needs and specific requirements for security, trust, or accountability. A recent publication [13] provides Governance model for AI in Healthcare (GMAIH) that covers the introduction and implementation of AI models in health care. This model includes recent requirements from United States, the Food and Drug Administration (FDA) [14] institute about requirements for AI systems.

From this point of view, clarity is a very important part of the creation and deployment of AI systems in the real environment and should be included in the design of the AI system. Of course, there are exceptions in this area as well, applications in which explainability does not play such an important role, especially business applications that focus on model accuracy and the potential profit and for which time

devoted to a deeper understanding of models would be cost-inefficient.

C. Desiderata of XAI models

In the literature on explainability, we often come across the term "desiderata", which we could translate as necessary requirements for XAI methods. These requirements represent aspects or properties that are expected and required from a method capable of explaining AI models. These requirements also vary in the literature or are intended for specific types of methods e.g. Desiderata for gradient methods [4] or Desiderata for interpretable model [15].

General requirements to be met by XAI models include also fidelity or honesty [16] [17]. Other requirements include robustness or stability, which measures whether similar input instances generate similar conclusions [18] as well as interpretability or comprehensibility [19], which means measures how difficult is for a person to understand the results from a given XAI model.

Other requirements, that were defined in [20] for XAI methods are Expressive Power, Translucency, Portability, or Algorithmic Complexity. For individual explainability, authors defined other necessary properties such as accuracy, fidelity, consistency, comprehensibility, certainty, degree of importance, novelty, and representativeness.

However, all of these desiderata depend on the specific application or environment in which the models will be deployed. The authors of the article on the Deployment of Explainable Models [21] argue that these requirements should be designed only based on the selected application and environment. It should be based on the following three points: 1. Identify stakeholders, 2. The involvement of each of the stakeholders, 3. Understand the reasons for explanation.

Based on these requirements, it is possible to compare models and select those that are suitable for the application we need e.g. in medicine [22]. However, recent practical approaches have shown that this comparison may not be sufficient and that more attention needs to be paid to practice tests along with evaluations from domain experts using these models [23].

It is also possible to compare explanatory methods from the point of view of several levels. Authors in [7] propose three main levels for the evaluation of interpretability:

- Application level evaluation (real task): Implementation of models for explainability in a specific application and its testing on a real task. For example, software that will detect fracture sites based on X-ray records. The doctor could evaluate the quality of the explainability that the software offers to explain its intentions.
- Human level evaluation (simple task): This level of explainability is also within applications, but the evaluation quality is not performed by experts, but by ordinary people - testers who are cheaper and also choose explanations according to how they help them understand at their level.
- Function level evaluation (proxy task): This level does not require people. It is suitable if the class of methods that will be used and with which the target class can work, e.g. decision tree. This model can be bounded to achieve a better level of explainability, such as the decision tree pruning method.

D. XAI Methods

Due to the growing number of methods in the field of explainability, it is difficult to understand the advantages, disadvantages, or competitive advantages in different domains. Besides, there are a large number of variations or categories of XAI methods such as global methods (explain the behavior of the model on the whole data set), local methods (explain the prediction or decision on a specific example), ante-hoc (model is created in the AI training phase), post-hoc (the model is created only on trained models), surrogate (AI model is replaced by an interpretable model) or a directly interpretable model (decision trees or decision rules) is used.

Molnar in his book [24] generally categorize XAI methods into 3 types: (1) methods with internal interpretation, (2) model agnostic methods, and (3) example-based explanation methods.

(1) One of the easiest ways to get explainability for the AI model is to use an internally interpretable model such as linear regression, logistic regression, decision trees, or decision rules. This avoids the use of "black box" models. This natural explainability of models is often at the cost of model accuracy.

(2) Model agnostic methods separate the explanations from the machine learning model. This brings certain advantages [16] and a great advantage over model-specific methods is their flexibility. Agnostic methods can be used for a wide range of machine learning models such as ensemble methods or deep neural networks. Even the output of an XAI method, whether it is a graphical or user interface, also becomes independent of the machine learning model used. Usually, several types of models are used for machine learning problems, and agnostic explanatory methods can also help in comparing the results of individual machine learning models. These methods include, for example Partial Dependence Plot (PDP), Accumulated Local Effects (ALE) Plot, Feature Interaction (FI), Local Surrogate (LIME), or SHAPley Additive exPlanations. The last mentioned method provides very interesting possibilities of explanation which are understandable and well understood.

(3) Example-based XAI methods [24] are based on the selection of a specific set of data on which they explain the behavior of a machine learning model or explain the distribution of data. These methods are also model-agnostic because they make any model of machine learning more interpretable. The difference from the previous type of methods is that example-based methods explain the model by selecting a set of examples from the data, not by creating associations or by the importance of properties. These explanations only make sense if we can represent an instance of the examples in a way that is understandable to humans. This approach works well for data types that contain multiple contexts, meaning that the data has an image or text structure. It is more challenging to represent tabular data with such methods because an instance can consist of hundreds to thousands of less structured functions. It is better to describe an instance if there are only a few functions or there is a way to shorten the instance description to fewer functions. These methods help to understand the complex distribution of data.

However, there are methods that do not belong to any of the previous groups. We can define this group of XAI methods as model-specific explanations. In this group are several XAI methods working with specific deep learning models e.g. CNN, LTSM, or GAN for image processing [25] or video

processing models [26].

On the other hand, in addition to AI systems based on machine learning, there are also non-machine learning applications such as Knowledge-based systems, Semantic Web applications, or Cognitive assistants that also need an explanation. The recent study [27] provides a good overview of XAI methods for this kind of AI systems.

III. Summary of solved and open research problems

Evaluating XAI is one of the open challenges. There are two primary goals of evaluating XAI methods, namely the complexity-based model and the human study-based model. The size of the model is often used in the literature as a factor that measures interpretability. It is also necessary to add other properties to the character of the comparison. There are studies where models are evaluated on the basis of requirements of desiderata as mentioned earlier in this work.

There is a lack of studies, that would compare XAI methods on real problems and evaluating these methods by users or domain experts. A recent study [23] showed that there are differences in comprehending XAI methods in explaining the behavior of AI models.

In addition to these primary objectives, the challenges include an overall evaluation of the models as described in this study [4]. At present, we still suffer from the problem of choosing the right methods for the right application as well as determining whether the technologies will be deployed.

A recent article [8], describes several challenges in AI applications and the need for clarity for those applications, while also offering possible approaches to answer these questions. For example, Computer Vision uses deep neural networks and complex architectures. There are also various tasks such as semantic segmentation, object detection, scene reconstruction, etc. The main challenge of clarity in this context is to identify the pixels, or groups of pixels, responsible for detecting patterns, uncertainties, or errors. Explainability is often a form of visual inspection, as it is a natural way when we process visual data. Several XAI methods are designed to explain such neural networks - mostly CNN. This includes, for example, Saliency maps [28] or SmoothGrad [29]. The disadvantage of methods for the explainability of visual data is that they are often local and explain only one selected example, after individual layers of NN, and do not capture semantics. On the contrary, the advantage is the comprehensibility of the outputs of these methods. An interesting approach that could solve these problems is a method called the knowledge graph [8].

In healthcare, there are also several problems when using machine learning technologies together with XAI methods. In addition to the problem of safety and security in data processing, it is necessary that the deployed applications have the trust of doctors as well as patients. Therefore, these applications need to provide appropriate explanations. This includes creating appropriate user interfaces for physicians. These challenges associated with assertiveness in medical applications remain open to further research [30].

There are also other problems in the medical field that need to be addressed in the context of XAI. For example, User-Centric Explanations [22], where it is necessary to monitor the requirements of future users, in the design of systems that will include the explainability of AI decisions in such a way that these users are able to understand the output that this system offers. It is also an important issue in terms of explainability on the performance and transparency of the models used. Who will decide whether to move towards better accuracy of models and poorer explainability? At which level should the model be transparent, so that it cannot be misused? [22]. In some applications, it may have better accuracy at the top, while in others it may have transparency combined with explainability. These issues need to be addressed with clinical experts or physicians.

The authors of [31] addressed the issue of transparency as well as the depth of this transparency. In this work, they explain the problems with transparency, which can lead to security problems, as well as a poorer performance of the model. They provided several types of transparency, the view of the end-user and the developer. Transparency may not always be a universal good. The challenge in creating systems is to find a suitable compromise between model performance, clarity, and transparency.

IV. Conclusion and Future direction

The use of artificial intelligence in healthcare can be of great benefit for the improvement of healthcare. However, it is first necessary to build the trust of doctors and patients in using AI technology. AI explanation methods can help us to improve trust in these technologies. In addition to trust, these methods can also serve as data-driven support for doctors' decision-making in important decisions. Future research will focus on user-centric explanations in the medical domain. These explanations should aim to foster trust and increase clinicians' understanding of AI system. User-centric explanations should be based on specific use cases and problem domains but also target users' knowledge and comprehension skills. That can be a way how to fill the gap between the users' reasoning processes and the ways we expect XAI methods to reason.

Acknowledgment

The work was supported by The Slovak Research and Development Agency under grants no. APVV-16-0213 and APVV-17-0550.

References

[1] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 361–370, 08 2016. [Online]. Available: https://doi.org/10.1093/jamia/ocw112

[2] M. A. Ahmad, A. Teredesai, and C. Eckert, "Interpretable machine learning in healthcare," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 447–447.

[3] G. Bologna and Y. Hayashi, "Characterization of symbolic rules embedded in deep DIMLP networks: A challenge to transparency of deep learning," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 7, no. 4, pp. 265–286, 2017.

[4] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey," *arXiv*, jun 2020. [Online]. Available: http://arxiv.org/abs/2006.11371

[5] G. Montavon, W. Samek, and K. R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing: A Review Journal*, vol. 73, pp. 1–15, 2018.

[6] L. Edwards and M. Veale, *Slave to the Algorithm? Why a Right to Explanationn is Probably Not the Remedy You are Looking for*. School of Law, Duke University, 2017, vol. 0.

[7] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv: Machine Learning*, 2017.

[8] F. Lecue, "On the role of knowledge graphs in explainable AI," *Semantic Web*, vol. 11, no. 1, pp. 41–51, 2020.

[9] Z. C. Lipton, "The mythos of model interpretability," *Communications of the ACM*, vol. 61, no. 10, pp. 35–43, 2018.
[10] J. Healey, "ARTIFICIAL INTELLIGENCE About Artificial intelligence," ... , *Artificial Intelligence and Financial Services*, vol. 9860, p. 60, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-29761-9_6
[11] High-Level Independent Group on Artificial Intelligence (AI HLEG), "Ethics Guidelines for Trustworthy AI," *European Commission*, pp. 1–39, 2019.
[12] S. Pichai, "AI at Google: our principles," *2018-07-07*, p. 1, 2018. [Online]. Available: https://blog.google/technology/ai/ai-principles/
[13] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 491–497, 2020.
[14] U. F. Administration and Drug, "Deciding When to Submit a 510(k) for a Change to an Existing Device Draft Guidance for Industry and," vol. 510, pp. 1–77, 2016.
[15] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, aug 2018.
[16] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
[17] G. Plumb, D. Molitor, and A. Talwalkar, "Model agnostic supervised local explanations," 2018.
[18] D. Alvarez-Melis and T. S. Jaakkola, "On the robustness of interpretability methods," 2018.
[19] M. Narayanan, E. Chen, J. He, B. Kim, S. Gershman, and F. Doshi-Velez, "How do humans understand explanations from machine learning systems? An evaluation of the human-interpretability of explanation," 2018.
[20] M. Robnik-Šikonja and M. Bohanec, *Perturbation-Based Explanations of Prediction Models*. Springer International Publishing, 2018. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-90403-0{_}9
[21] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley, "Explainable machine learning in deployment," *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 648–657, 2020.
[22] M. A. Ahmad, A. Teredesai, and C. Eckert, "Interpretable machine learning in healthcare," *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, p. 447, 2018.
[23] S. Jesus, C. Belém, V. Balayan, J. a. Bento, P. Saleiro, P. Bizarro, and J. a. Gama, "How can i choose an explainer? an application-grounded evaluation of post-hoc explanations," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 805–815. [Online]. Available: https://doi.org/10.1145/3442188.3445941
[24] C. Molnar, *Interpretable Machine Learning*, 2019, https://christophm.github.io/interpretable-ml-book/.
[25] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Explainable COVID-19 detection using chest CT scans and deep learning," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–22, 2021. [Online]. Available: http://arxiv.org/abs/2011.05317
[26] D. R. Chittajallu, B. Dong, P. Tunison, R. Collins, K. Wells, J. Fleshman, G. Sankaranarayanan, S. Schwaitzberg, L. Cavuoto, and A. Enquobahrie, "XAI-CBIR: Explainable ai system for content based retrieval of video frames from minimally invasive surgery videos," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, no. Isbi, pp. 66–69, 2019.
[27] S. Chari, D. M. Gruen, O. Seneviratne, and D. L. McGuinness, "Foundations of explainable knowledge-enabled systems," *arXiv*, no. March, 2020.
[28] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, pp. 1–8, 2014.
[29] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "SmoothGrad: Removing noise by adding noise," *arXiv*, 2017.
[30] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable AI in Healthcare," *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, Cyber SA 2020*, no. June, 2020.
[31] A. Weller, "Transparency: Motivations and Challenges," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019.

Analyzing and identifying potential factors most significantly affecting CVD

[1]*Zuzana PELLA (4th year)*
Supervisor: [2]Ján PARALIČ

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]zuzana.pella@tuke.sk, [2]jan.paralic@tuke.sk

***Abstract*— This paper is motivated with the actual need to deal with early diagnosis of cardiovascular diseases (CVD) in order to prevent them. A multidisciplinary project presented here is focused on the early diagnosis of CVD and understanding of a broader set of factors influencing CVD. The next part of the publication is devoted to the tasks using machine learning to fulfill the project. For this purpose, we focus on using classical methods of machine learning, such as decision trees or regression methods. We will enrich them with statistical tests and results of factor analysis.**

***Keywords*—data analysis, decision support system, factor analysis, machine learning.**

I. Introduction

As we mentioned in the previous publication [1], our dissertation thesis is a part of the multidiscipline project Kosice Selective Coronarography Multiple Risk Study (KSC MR Study). We cooperate with a medical specialists from the East Slovak Institute for Cardiovascular Diseases (ESICS) and the Pavol Jozef Šafárik University Medical faculty. The presented KSC MR study (method paper) establishes a novel approach in cardiovascular prevention-based not on statistical methods but more towards personalized medicine supported by machine learning [2]. The need to address this issue is the high mortality rate due to cardiovascular diseases (CVD). According to the 2017 Eurostat report, Slovakia was the 3rd most affected of the European Union countries. In Slovakia, the standardized mortality rate represented 3753 deaths per million inhabitants in 2016 [4]. The ratio of the number of hospitalizations is relatively high. In 2014 – 2018, the hospitalization rate due to circulatory system diseases represented 15.2% of all hospitalizations in Slovakia. Of which up to 55.5% represented hospitalizations due to ischemic heart disease (IHD, medical designation I20 – I25) and other heart diseases (medical designation I30 – I52) [5]. Like KSC MR Study, our dissertation thesis considers this topic significant and essential for improving our citizens' quality of life. We believe that by combining machine learning methods with available patient data, we can capture the disease's onset in time, ultimately improving the quality of life and relieving our health system.

II. The initial state of our research

We use data to reach our goals, which originates from patients hospitalized in ESICS [1]. The form of raw data is a semi-structured document, a medical record. The first step of our project participation was the proposal and development of a software application for text processing [3]. The Pals application transforms incoming medical records into structured table data. This data's structure and meaning are explained in our previous publications [1][3].

We used this retrospective data for several types of research studies. For example, the aim of the one described in [6] was based on selection various groups of attributes (the selection of attributes was the result of the applied method Forward Stepwise Selection focusing on criteria: Adjusted RSq, Mallow's Cp, Bayes information criterium BIC) to determine the need for medical examination of selective coronarography. Unfortunately, the accuracy of the created model was not satisfactory.

III. Current issues

For the last months, our work has been devoted to two parallel tasks:

1. The analysis of the impact of factors on early detection of cardiovascular risk.
2. Factor analysis and its use for determination of cardiovascular risk.

Both these tasks work with data that contains 808 records (each patient has only one record). This data originates from ESICS, where patients were hospitalized between June 2017 and March 2018. Throughout the development phase, we follow the CRISP-DM methodology. We should merge the results of both of these complex tasks to one final decision support system at the end.

1. The analysis of the impact of factors on early detection of cardiovascular risk

2. The first more fundamental issue we needed to solve in our dataset was many missing values. We decided to use a very known method for imputation of missing values – k Nearest Neighbors (kNN). The fundamentals of kNN imputation are to find k similar complete instances to the incomplete ones and to add missing values based on voting [10]. Also, we decided to use one more method, known as Multivariate imputation by chained equations (MICE) [7]. MICE is based on the correlation between variables. The process of filling missing values goes on through several iterations using various methods, such as random forest algorithm, predictive mean matching, classification, and regression trees algorithm [8]. On the base selected method, each iteration of MICE creates a new

model filling out missing values concerning maintaining relationships in the data and preserving the uncertainty about these relations [9].

In the next phase, we tried to create four groups of patients (fluently from the KSC MR Study project specification):

1. Patients with advanced atherosclerosis, severe results of coronary angiography examination.
2. Patients in the middle age and positive coronarography findings or patients with a high CVD risk score and positive coronarography findings.
3. Patients with a high CVD risk score but good results from the coronary angiography examination.
4. Healthy persons.

That division of patients was a part KSC MR Study proposal. Also, we were interested in whether the division into groups influences the need to perform selective coronarography. We try to create several models – base on the CART algorithm, Random Forest (RF), Naïve Bayes, and SVM. Unfortunately, the results of our experiment are not satisfactory. For example, we achieved accuracy between 36,92% (CART with kNN imputation) and 49,26% (RF with kNN imputation). However, we can say that kNN imputation has slightly better results (for comparison, 42,96% (RF with MICE imputation) vs. 49,26% (FR with kNN imputation)).

This part of our interest, like our overall research, was consulted with the medical expert. We presented the results of the usage of individual attributes within the created models. This usage (understood as the significance of attributes) corresponds to medical practice regarding the influence of monitored CVD factors. Specifically, these attributes describe ECG rhythm, (non)drinking alcohol, and IHD in the patient's history.

2. Factor analysis

Factor analysis (FA) is a statistical method. FA is not a reduction method but a summarized method that finds and ranks essential factors. The result of its usage is grouping attributes into a smaller number of factors [11].

Like in the previous section, for FA, it was also necessary the imputation of missing values. We used kNN imputation. The next step was handling the outlier for attributes Vyska, Vaha, and BMI value. Subsequently, it was necessary to determine that our data are suitable for the FA application based on correlation analysis (needed over 0.5) and measure of sampling adequacy (MSA, over 0.5). Inappropriate or unusable attributes have been removed - the attribute did not meet the required criteria or contained too many missing values. We focused on two types of experiments: the first with all available attributes, including the target; the second without the target, and 10 describing the result of the coronarography.

The first experiment was designed to find context between the attributes describing the result of the coronarography and all others. We created a scree plot to estimate the number of factors. The slope of the graph curve suggests the most suitable number of factors. For our case, these are 3-4 factors. Using the oblimin rotation, we performed FA, and we present the result for the extraction of 3 factors listing particular attributes assigned to each of the identified factors:

1. *Factor*: *Nalez* (degree of obligation to find coronarography); *RIA* (percentage narrowing of the RIA branch); *ACD* (percentage narrowing of the ACD branch); *RCX* (percentage narrowing of the RCX branch); *O_IM* (the occurrence of a heart attack in the patient's history); *O_ISCH* (the occurrence of IHD in the patient's history); *ACS* (percentage narrowing of the ACS branch); *EKG_STE* (the presence of elevations in the ST region);
2. *Factor*: *Pohlavie* (Gender); *Vyska* (Height); *Fajcenie* (Smoking); *F_Pocet* (number of smoked cigarettes per day); *Alkohol* (alcohol consumption);
3. *Factor*: *Vaha* (Weigh); *Urea* (blood urea level); *BMI* (body mass index); *ECHO_EF* (left ventricular ejection fraction - %); FBG (fibrinogen levels); *ECHO_PH* (degree of pulmonary hypertension); R_*ICHS* (occurrence of IHD in the family); *EKG_Rytmus* (occurrence of IHD in the family); *R_IM* (occurrence of a family infarction).

This result was presented to the medical specialist. From the physician's point of view, the representation of factors helps control specific patients' inclusion in the study.

The second experiment was designed to reduce the number of attributes for creating the models. Based on the recommendation the slope of a scree plot, we applied FA to create three factors:

1. *Factor*: *Pohlavie* (Gender); *Vyska* (Height); *Fajcenie* (Smoking); *F_Pocet* (number of smoked cigarettes per day); *Alkohol* (alcohol consumption); *Rocnik* (year of birth); *F_Doba* (number of years of smoking / non-smoking);
2. *Factor*: *O_IM* (the occurrence of a heart attack in the patient's past); *O_ISCH* (the occurrence of IHD in the patient's history); *O_DM* (occurrence of type 2 diabetes in the patient's history); *Vaha* (Weigh); BMI (body mass index); *O_CMP* (occurrence of stroke in the patient's history);
3. *Factor*: *R_ICHS* (occurrence of IHD in the family); *R_IM* (occurrence of a family infarction); *Urea* (blood urea level).

Reducing the number of attributes in this experiment may help find a potential link between the attributes that have not been considered significant to CVD. Also, reducing this way, using FA, allows us to reduce the number of attributes, but we do not lose essential information.

III. Future work

In the next phase of our work, using regression methods, statistical tests, and FA, we will identify potential factors most significantly affecting CVD. Based on them, models will be built to fulfill the dissertation's goals with the highest possible accuracy and the KSC MR Study. The result of our work will be processed as a DSS. It should be applicable in the early diagnosis of coronary heart disease and an effort to determine the need for selective coronary angiography. The whole process will be regularly consulted with a specialist doctor.

Acknowledgment

The work was supported by The Slovak Research and Development Agency under grant no. APVV-17-0550.

References

[1] Pella, Z.: Data mining research for reduction of cardiovascular diseases risks. In: SCYR 2018: Proceedings from Conference. Košice: TU, 2018, pp. 75-76. ISBN 978-80-553-2972-7

[2] Pella, D., Tóth, Š., Paralič, J., et al.: The possible role of machine learning in detection of increased cardiovascular risk patients – KSC MR Study (design). In: Archives of Medical Science. 2020. doi:10.5114/aoms.2020.99156.

[3] Pella, Z., Milkovič, P., Paralič, J.: Application for text processing of cardiology medical records. In: Proceedings of the IEEE World Symposium on Digital Intelligence for Systems and Machines (DISA 2018), pp. 169-174, IEEE (2020).

[4] Eurostat: Your key to European statistic: Deaths due to coronary heart diseases in the EU. 28.9.2020 [online: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20200928-1].

[5] Národné centrum zdravotníckych informácií: Kardiológia: Hospitalizácie na choroby obehovej sústavy v Slovenskej republike za roky 2014 – 2018. 17.9.2019 [online: http://data.nczisk.sk/statisticke_vystupy/Kardiologia/Hospitalizacie_na_choroby_obehovej_sustavy_v_SR_2014_2018.xlsx].

[6] Pella, Z.; Lohaj, O.; Paralič, J.; Pella, D. Jr.: Analýza významnosti faktorov vplývajúcich na závažnosť kardiovaskulárnych ochorení. In: Data a Znalosti & WIKT 2019. Košice: TU, 10.-11.10.2019, pp. 16-19. ISBN 978-80-553-3354-0.

[7] Azur, M. J.; Stuart, E. A.; Frangakis, C.; Leaf, P. J.: Multiple imputation by chained equations: what is it and how does it work?. In: International Journal of Methods in Psychiatric Research, vol. 20, no. 1. 2011, pp. 40-49. doi: 10.1002/mpr.329.

[8] RDocumentation: mice: mice: Multivariate Imputation by Chained Equations [online: https://www.rdocumentation.org/packages/mice/versions/3.13.0/topics/mice].

[9] van Buuren, S.; Groothuis-Oudshoorn, K.: mice: Multivariate Imputation by Chained Equations in R. In: Journal of Statistical Software, vol. 45, no. 3. 2011, pp. 67. doi: http://dx.doi.org/10.18637/jss.v045.i03.

[10] Zhang, S.: Nearest neighbor selection for iteratively kNN imputation.In: Journal of Systems and Software, vol. 85, no. 11. 2012, pp. 2541-2552. doi: https://doi.org/10.1016/j.jss.2012.05.073.

[11] Wang, H.-F.; Kuo, Ch.-Y.: Factor analysis in data mining. In: Computers & Mathematics with Applications, vol. 48, no. 10-11. 2004, pp.1765-1778. doi: https://doi.org/10.1016/j.camwa.2003.05.016.

Intelligent Methods Using Data to Support Decision Making

Viera PAVLIŠKOVÁ (1st year)
Supervisor: František BABIČ

Department of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

viera.pavliskova@tuke.sk, frantisek.babic@tuke.sk

***Abstract*— Let's start with the definition of what decision making in the medical domain is (Meng, Chapman, 2020): "it includes decisions in medical, policy, or everyday settings that impact individual and public health. It frequently involves risk and uncertainty, difficult trade-offs, information overload, decisions about future outcomes, and interdependent behaviours, as well as reliance on numbers." With the rapid development of information and communication technologies, the medicine and healthcare domain understood the richness's collected data. This finding was followed by the first attempts to cooperate with data analytics to exploit the potential of intelligent methods for data processing and analysis. This cooperation has shown that doctors or medical experts are now always able to consume analytical methods' decision models properly. In this paper, we investigate existing approaches with their advantages or disadvantages and based on the results, and we will propose an innovative visualisation method as an integral part of the clinical decision support system.**

***Keywords*— decision making, data, visualisation, understandability.**

I. Introduction

Rashidi et al. state that today's decision-making is much more complex than in the past for two crucial reasons [1]. First, the growing number of technologies and communication systems has led to many possible alternatives from which the decision-maker can choose. Second, the increased level of today's problems' structural complexity may result in a series reaction of increasing costs if an error can occur.

A decision support system (DSS) is generally used to model human reasoning and decision-making. Therefore, it can accept users' facts, process them, and propose solutions like domain experts. Consequently, the DSS can be very useful in supporting the evaluation of various decisions to select the strongest and cost-effective responses transparently and systematically.

Finding effective cooperation between a data analyst and physicians is often challenging and more complex than the task of extracting knowledge from visual models generated by computer systems. Rokošná et al. focused on effective communication between the doctor and the data analyst to ensure effective, fast, and exciting feedback of the proposed visualisations [2].

Based on the state-of-the-art analysis, we found that the decision-making models' performance and the form of its visualisation for end-users are important. The correct visualisation can improve its readability, understandability, and deployment in practice.

These facts motivated us to research new visualisation methods for decision models generated by selected data analytics algorithms. In other words, our task is to explain to the end-users the analytical models in a more familiar form for them.

II. Decision making in medicine

Nowadays, many changes are continually taking place in the field of medicine. Physicians strive to make the best possible decisions when making diagnoses, designing treatments, preventing them, or drawing conclusions. They are continually looking for new solutions to improve and confirm their decisions with confidential systems. Therefore, it is necessary to support medical decision-making regarding the analysis and prediction of a patient's condition, whether in treatment or diagnosis. The completeness and timeliness of the available patient information are essential when making decisions. To facilitate, speed up and streamline medical decisions, intelligent devices, intelligent systems, and methods are often integrated into the medical field [3][4].

The physicians are increasingly motivated to collaborate with data analysts who seek to develop intelligent systems for qualified medical care decisions. These systems can treat a particular patient who needs attention earlier, who is more likely to be hospitalized, or even delay eliminating delayed medical care [4]. Such intelligent systems will be responsible for collecting, processing, and forwarding data using medical records. Data analytics (DA) can solve various tasks such as data pre-processing, improving the accuracy of predicting a patient's disease's development, the exact factors that lead to determining the required treatment, and others [5][6].

Bhanot et al. [7] emphasize that visualization plays a central role as an interpretable medicine tool. They presented some open-ended questions like a selection of the most appropriate visualization techniques for heterogeneous and structured data that commonly occur in the biomedical context; an assessment of the essential elements to be visualized to facilitate the interpretation of real experts or to integrate the requirements/limitations of domain experts into the ML model and various others.

Reyes et al. [8] pointed out that interpretation methods can improve the understanding, trust, and verification of AI radiological systems, and increasing the interpretability of the system will also increase trust from the professional user. Thus, interpretability will support effective acceptance in practice.

The authors [9] claim that the resulting framework must focus not only on machine learning (ML) models and visual techniques developed by the analyst but also on medical and healthcare professionals who provide expert validation and in-depth analysis of the results obtained.

III. Medical Data Analytics

Currently, data analytics models are used in many areas, like healthcare and medicine [10]. Information and communication technologies move forward every year, so there is a constant need to find new, unknown, valuable, and hidden information from medical records. Medical mining data focuses on diagnosing and predicting diseases, detecting insurance fraud in health insurance, identifying effective treatment methods, treating treatment effectiveness, improving services provided to patients, and increasing hospitals' evaluation.

A. Typical Decision Models

Over the years, many visualisation methods have been developed to represent a large amount of information and examine it. These methods include usability, interactivity, interface properties, and others used to visualise data because they have an evaluation mechanism. There are many ways to visualise data, but it depends on the scientist which method of data representation and presentation he chooses [11].

Decision trees represent the most popular way how to solve the classification tasks because the extracted knowledge has a simple, intuitive, and understandable interpretation [12][13]. They offer end users a clear graphical form of displaying the results and provide a reasonable interpretation for a quick and straightforward evaluation of the acquired knowledge. The typical visualisation consists of nodes and edges (see Fig. 1).

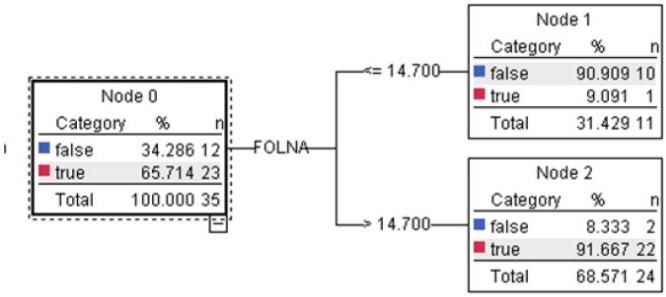


Fig. 1 Example of decision tree generated from medical data sample [14]

Tolles et al. describe logistic regression as a standard statistical method for quantifying the relationship between patient characteristics and clinical outcomes [14]. Liu applied logistic regression to classify the data set of breast cancer [16].

The purpose of clustering algorithms is to identify groups of objects, or clusters, that are more like to each other than to another cluster [16]. In general, this approach focuses on extracting a set of variables providing a complex explanation about relevant aspects of a data sample. Yeboah et al. aimed to heterogeneous clinical data, specifically traumatic brain injury [18]. They applied three different combinations of algorithms: a combination of k-means, spectral and agglomerative algorithms; a combination of k-means, spectral, agglomeration and gaussian algorithms; an aggregation of spectral, agglomerative, and Gaussian algorithms. The two best results of these models were visualized using the ISOMAP algorithm and a canonical plot. ISOMAP aims to discover a low-dimensional embedding of high-dimensional data by maintaining the geodesic distances between all points [19].

Association rules (AR) mining is a method to uncover how items are associated with each other [20]. AR are usually represented in the form of the implication X => Y, where X and Y are disjoint sets of items. Rule quality is usually measured by rule support and confident. Huang et al. discovered the association rules of chronic diseases from massive data, which are constantly collected through health examinations and medical treatments [21].

B. New Visualisations

The Sunburst diagram is a new, intuitive way of displaying a decision tree visualising hierarchical data radially. The decision tree is visualised as if we were looking at the top-down decision tree, not as traditionally from the side. Sunburst may resemble nested pie charts, with the top of the hierarchy, the root of the decision tree being in the middle, and the deeper levels tied to that centre. The diagram consists of rings that represent the individual levels of the decision tree. These rings are composed of arcs, each arc being characterised by a specific attribute, based on which the decision tree is further branched. Arcs represent nodes of the original decision tree, and the length of the individual arcs correspond to the percentage of the given attribute depending on the total number of examples in the dataset [22].

After a detailed analysis of existing solutions, we found that this method is not currently sufficiently researched in the field of medicine. We present a work of authors analysing the primary sources of user interface framing for e-medical health records and used Sunburst visualisation techniques to understand the essence of EHR fragmentation, optimisation, and redesign [23]. In this study, 2 Sunburst graphs were created for two different approaches to the EHR user interface. The results show that the degree of fragmentation for the first conventional commercial EHR is relatively high (see Fig. 2) and could probably lead doctors to spend much time navigating, viewing, and storing information. In the second EHR, the authors found that the degree of fragmentation is lower and contains fewer levels.

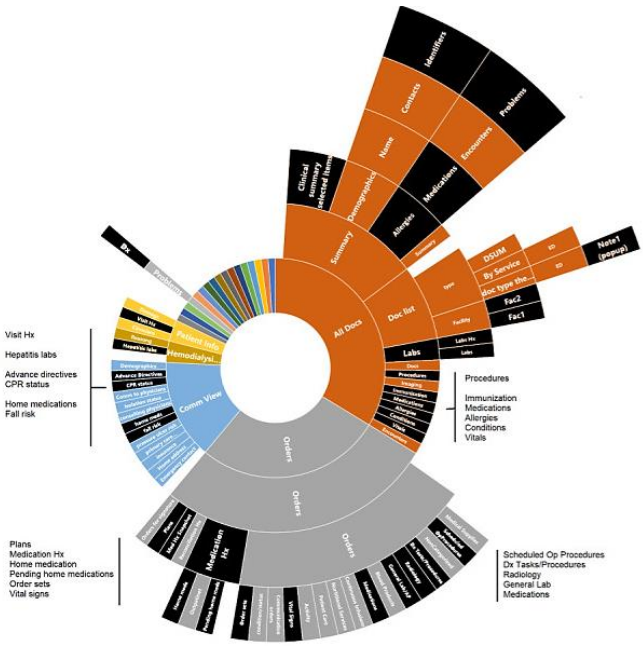


Fig. 2 Sunburst diagram representing EHR with a higher degree of fragmentation [23].

The SHapley Additive exPlanation (SHAP) values are developed by Shapley in cooperative game theory [24]. SHAP is an acronym for the method intended for predictive models, and to avoid confusion, we will use the term Shapley value. It provides an effective and thorough measure of the importance of elements in a model. Lundberg and Lee explored different methods and combinations of SHAP values [25]. They launched a Python package that can calculate Shapley values for various techniques such as LightGBM, GBoost, CatBoost, XGBoost, and tree models from the Scikit-learn library [26]. Koo et al. aimed to develop machine learning models such as Lasso, Ridge, Support Vector Machine, Random Forest and XGBoost to predict remission in patients treated with biological agents using data rheumatoid arthritis patients and identify essential properties having the most significant impact on biological agents [27]. They generated a Shapley graph interpreting the importance of the attributes for each biological substance, see Fig. 3 .

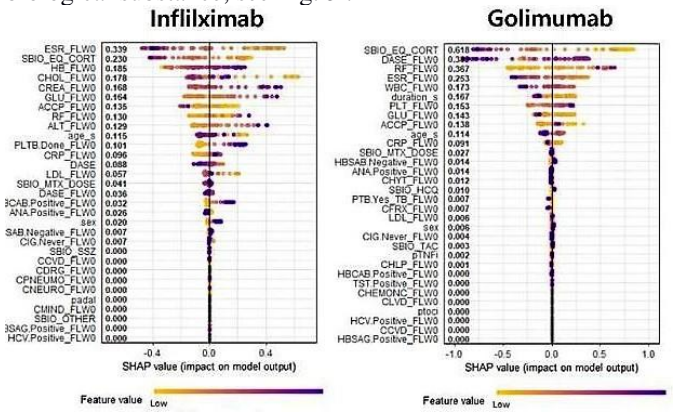


Fig. 3 Shapley plots and SHAP values for the feature importance [27]

Understanding analytical models' behaviour allows users to trust model predictions and make decisions based on those predictions. The models' interpretability has become a key concern of data analytics, machine learning or artificial intelligence. Ribeiro et al. began to explain machine learning models using model-agnostic approaches, which treat models as a black-box model [28]. They created the LIME (Local Interpretable Model-agnostic Explanations) method, which is used to identify the interpretable model and explain any classifiers. Li et al. used the LIME technique to explain the results of decision trees [29]. One of the main research challenges was how to improve interpretability while preventing deterioration in model accuracy. They proposed a modified LIME algorithm based on a locally adapted regression of the decision tree called tree-LIME, which is a model-agnostic method. The XGBoost algorithm was used to create a tree-LIME model, see Fig. 4.

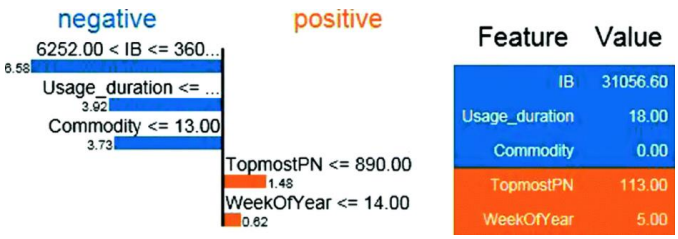


Fig. 4 Representation of LIME [29]

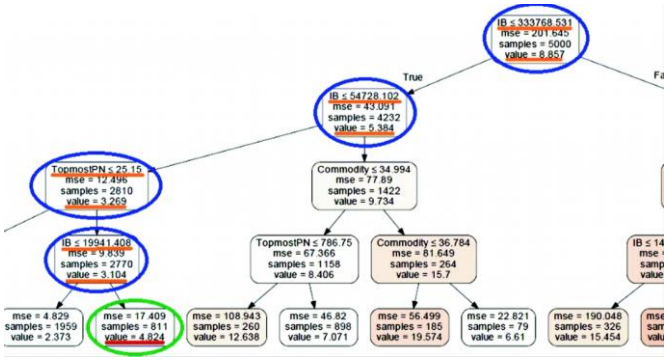


Fig. 5 Representation of tree-LIME [29]

Guo et al. aimed to analyse and demonstrate the electrocardiographic (ECG) left ventricular hypertrophy (LVH) with ST-segment elevation (STE) data using a heat map. They investigated the feasibility and clinical value of heat mapping for ECG data visualisation [30]. The authors found that heat maps and cluster analysis could be used to fully visualise each lead of an electrocardiogram and provide relatively complex information.

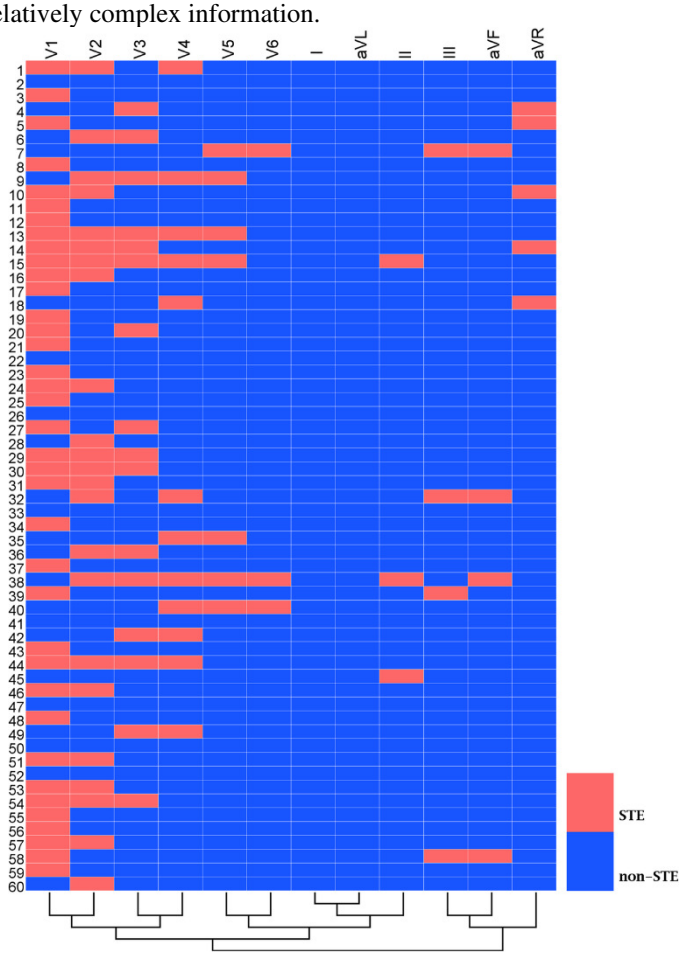


Fig. 6 Heat map corresponding to an ST-segment elevation amplitude that reached the threshold specified in the STEMI guidelines [30]

IV. Further Research

This paper points out the importance of the understandability of data analytics models in the medical decision-making process. Explaining individual models' decisions for the doctor is currently very desirable due to the efficient and simple work with the resulting models. Each of these studies focuses on a separate medicine area and analyses a different population sample. To ensure greater efficiency and understandability of the model, intensive cooperation

between data analysts and relevant experts in particular areas is essential.

In our dissertation, we focus on analysing available medical samples to gain potential new and valuable knowledge. In co-design with medical experts, we will examine the traditional methods used for decision models visualisation and propose new possibilities of model visualisation to improve acquiring a higher understanding and applicability of decision models.

ACKNOWLEDGMENT

The work was supported by The Slovak Research and Development Agency under grants no. APVV-16-0213 and APVV-17-0550.

REFERENCES

[1] M. Rashidi, M. Ghodrat, B. Samali, M. Mohammadi. "Decision Support System," in *Management of Information Systems,* BoD – Books on Demand, 2018.

[2] J. Rokošná, F. Babič, L. T. Majnarić, L. Pusztová, "Cooperation Between Data Analysts and Medical Experts: A Case Study," presented at the 4th IFIP TC 5, Dublin, Ireland, Aug 25-28, 2020.

[3] J. Waring, C. Lindvall, R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine*, vol. 104, Apr. 2020, doi: https://doi.org/10.1016/j.artmed.2020.101822.

[4] M. Pourhomayoun, M. Shakibi, "Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making," *MedRxiv*, Apr. 2020, doi: https://doi.org/10.1101/2020.03.30.20047308.

[5] R. Tkachenko, I. Izonin, 2 "Model and Principles for the Implementation of Neural-Like Structures based on Geometric Data Transformations," presented at the Advances in Computer Science for Engineering and Education, Kiev, Ukraine, Jan. 18-20, 2018.

[6] R. Tkachenko, I. Izonin, K. Kryvinska, V. Chopyak, N. Lotoshynska, D. Danylyuk, "Piecewise-linear Approach for Medical Insurance Costs Prediction using SGTM Neural-Like Structure," presented at the the 1st International Workshop on Informatics & Data-Driven Medicine Lviv, Ukraine, Nov. 28–30, 2018.

[7] G. Bhanot, M. Biehl, T. Villmann, D. Zühlke, "Biomedical data analysis in translational research: Integration of expert knowledge and interpretable models," presented at the 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Burges, Belgium, Feb. 20, 2017.

[8] M. Reyes, et al. "On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities," Radiology: Artificial Intelligence, vol. 2, no. 3, May 2020, doi: https://doi.org/10.1148/ryai.2020190043.

[9] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural Comput & Applic, vol.* 32, pp. 18069–18083 Feb 2020, doi: https://doi.org/10.1007/s00521-019-04051-w.

[10] S. Ch. Pandey, "Data Mining Techniques for Medical Data: A Review," presented at the International Conference on Signal Processing, Communication, Power and Embedded System, Paralakhemundi, India, Jun 26, 2017.

[11] S. S. Ajibade, A. Adediran, "An Overview of Big Data Visualization Techniques in Data Mining," *International Journal of Computer Science and Information Technology Research,* vol. 4, no. 3, pp. 105-113, Jul. 2016.

[12] J. Paralič, "Objavovanie znalostí v databázach," Košice: Elfa, 2003, pp. 69 - 78.

[13] L. Rokach, M. Maimon, "Data Mining and Knowledge Discovery Handbook," 2nd ed. New York: Springer, 2010, pp. 26-29.

[14] F. Babič, L. Majnarić, A. Lukáčová, J. Paralič, A. Holzinger, "On Patient's Characteristics Extraction for Metabolic Syndrome Diagnosis: Predictive Modelling Based on Machine Learning," *In: Bursa M., Khuri S., Renda M.E. (eds) Information Technology in Bio- and Medical Informatics,* vol. 8649, 2014, doi: https://doi.org/10.1007/978-3-319-10265-8_11.

[15] J. Tolles, et al. "Logistic Regression Relating Patient Characteristics to Outcomes," *JAMA Guide to Statistics and Methods*, vol. 316, no. 5, Aug. 2016.

[16] L. Liu, "Research on Logistic Regression Algorithm of Breast Cancer Diagnose Data by Machine Learning," 2018 International Conference on Robots & Intelligent System (ICRIS), Changsha, China, 2018.

[17] M. Z. Rodriguez, C. H. Comi, D. Casanova, O. M. Bruno, D. R. Amancio et al. "Clustering algorithms: A comparative approach," PLOS ONE, vol. 14, no. 1, Jan. 2019, doi: https://doi.org/10.1371/journal.pone.0210236

[18] D. Yeboah, et al. "An Explainable and Statistically Validated Ensemble Clustering Model Applied to the Identification of Traumatic Brain Injury Subgroups," IEEE Access, vol. 8, oct. 2020, doi: 10.1109/ACCESS.2020.3027453.

[19] J. B. Tenenbaum, V. Silva, J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science,* vol. 290, 2000, doi: 10.1126/science.290.5500.2319.

[20] A. Capozzoli, T. Cerquitelli, M. S. Piscitelli, "Pervasive Computing: Next Generation Platforms for Intelligent Data Collection," Elsevier, Inc., 2016.

[21] M. Huang, H. Sung, T. Hsieh, M. Wu, S. Chung, "Applying data-mining techniques for discovering association rules," *Soft Computing,* vol. 24, Jul. 2019.

[22] "A New Way to Visualize Decision Trees," available at: https://blog.bigml.com/2013/04/19/a-new-way-to-visualize-decision-trees/, April 19, 2013.

[23] Y. Senathirajah, et al. "Characterizing and Visualizing Display and Task Fragmentation in the Electronic Health Record: Mixed Methods Design," *JMIR Publications*, vol. 7, no. 4, Oct. 2020, doi: doi:10.2196/18484.

[24] S. L. Shapley, "A Value for n-Person Games," In *Contributions to the Theory of Games,* edited by Harold W. Kuhn and Albert W. Tucker, 307–17. Princeton: Princeton University Press, 1953.

[25] S. M. Lundberg, S. Lee, "A Unified Approach to Interpreting Model Predictions," presented at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, Dec. 04-09, 2017.

[26] S. M. Lundberg G. G. Erion, S. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," *Computer Science*, 2018, doi: https://arxiv.org/abs/1802.03888.

[27] B. S. Koo, et al. "Prediction of Remission for each Biologics based on Patient's Clinical Information Before Starting Biologics Using Explainable Artificial Intelligence: Data From the Korean College of Rheumatology Biologics Registry," *Annals of the Rheumatic Diseases,* pp. 17-18, doi: http://dx.doi.org/10.1136/annrheumdis-2020-eular.3438.

[28] M. T. Ribeiro, S. Singh, C. Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," presented at the 22nd International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, Aug. 2016, pp. 1135-1144, doi: https://doi.org/10.1145/2939672.2939778.

[29] H. Li, W. Fan, S. Shi, Q. Chou, "A Modified LIME and Its Application to Explain Service Supply Chain Forecasting," presented at the 8th CCF International Conference on Natural Language Processing and Chinese Computing, Dunhuang, China Oct. 9-14, 2019, doi: https://doi.org/10.1007/978-3-030-32236-6_58.

[30] H. Guo, et al. "Heat map visualization fo electocardiogram data analysis," *BMC Cardiovasc Disord,* no. 277, Jun. 2020.

Rules Adaptation Mechanism in the Medical Diagnostic Process

Ľudmila PUSZTOVÁ (4th year)
Supervisor: František BABIČ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

ludmila.pusztova.2@tuke.sk, frantisek.babic@tuke.sk

***Abstract*—Decision model explainability is one of the major research topics nowadays. It means that the data analysts try to generate these models in a simple, understandable form for the end users. The success of this approach depends on the analytical method used for the model's generation. For example, the decision trees can be visualized in the form of a graph. The second aspect is the relevance of these models and their adaptation to the new conditions or new patients. A traditional procedure covers a re-generation of the whole rules set with each new patient. This procedure motivated us to design a semi-automatic adaptation mechanism in the case-based reasoning process as a part of the decision support system. We tested and validated our procedure in the medical diagnostic process. The experiments and obtained results were published in two international journals and conferences.**

***Keywords*—Case-based Reasoning, Decision Support System, Rule adaptation.**

I. Introduction

Explainable artificial intelligence (XAI) represents one of the hottest topics in the domain of data processing and analysis. AI systems have sophisticated decision support mechanisms based on an enormous amount of processed and efficiently synthesized data into complex models. Due to their internal coding and complexity level, many of these models can be described as black boxes, but they are not accessible to humans and cannot be interpreted. However, attention is currently focused on the transparency and explainability of such models.

With the growing availability of structured and unstructured data and rapid advances in analytical techniques, artificial intelligence (AI) is revolutionizing the medical field. However, with the increasingly necessary role of AI in healthcare, there are growing concerns about the lack of transparency and clarity and possible prejudices and doubts about the diagnoses obtained from various prediction models. XAI should increase the trust in these models.

Decision trees are one of the most used and understandable AI methods. They allow an explanation of the whole patient's diagnostical path. This path is covered from a root node to the leaf node, expressed in the form of generated rules. We can use these rules to create a decision support system. In this case, an interesting research question is – it is necessary to re-generate the whole list of rules with each new patient? Or we can use some adaptation mechanism.

A. Important terms definition

The work is primarily focused on **Case-based reasoning (CBR)** and the adaptation step of this method. CBR method simulates human thinking by reusing previous solutions based on similarity to a new case. This method traditionally provides finding the most similar cases to a new case and performing the rule adaptation step. The method consists of 4 steps – RETRIEVE (find and load the most similar cases), REUSE (reuse of the proposed solution), REVISE (control of the proposed solution), and RETAIN (save the learned case for future use). But the number of systems providing all these steps is very low. The adaptation step is often omitted because it is relatively complex [1][2].

Decision support systems (DSS) are designed to help qualified people use data, knowledge, documents, models, and communication to solve problems in the decision-making process [3]. The systems can receive facts from users, process them, and design solutions close to the domain expert´s point of view [4]. They are very popular in the medical field because they help doctors in everyday decision-making processes, such as diagnosing one or more diseases, designing appropriate treatment, diagnosing the absolute absence of a disease, or finding the cause of their occurrence [5].

II. Previous Work and Initial Status of the Research Task

Our main goal was to design a semi-automatic adaptation mechanism in the CBR process as a part of the decision support system. The proposed mechanism is a result of several iterations tested on available data samples. Its pre-final version was published at the CD-MAKE 2019 conference [6]. This mechanism offers the possibility to find the most similar patients from the database to a new patient (s), the new patient's classification, the rules adaptation, and storage of a new classified patient. It is a part of a medical support system and the implementation was realized in two steps.

Firstly, we created module 1 focusing on the creation of the decision trees models and selection of the most successful one. Inputs to this module are entered only by the data analyst. Module 2, representing the user interface, uses the most successful model and knowledge base filled with a set of sub-rules from module 1. The system offers the following functions for its users:

- rules and sub-rules generation from the decision model,
- write-in a new patient,
- a summary of historical patients in the database,

- finding the coverage of a new patient and historical patients by the rules,
- finding similarities between new and historical patients,
- display helping tables to support the expert's decision classification,
- storage of a classified patient,
- adaptation of sub-rules and their storage,
- generation of reports with important system outputs.

The detailed description of the mechanism and the system was published in the peer-reviewed journal Applied Sciences, Special Issue "Machine Learning in Medical Applications", Q2 rate by JCR [7].

We experimentally tested and validated our mechanism using four different medical data samples. We managed one case study with a cardiologist from the East Slovak Institute of Cardiovascular Diseases (VÚSCH) in Košice. The case study consisted of some model scenarios: the system found the same patient; the system found a very different patient; and the system found partially similar patients. He also tested the possibility of adapting the rules and their subsequent impact on the classification. We collected his feedback through a questionnaire, such as the decision support system was transparent, well arranged and designed, including all the crucial functions important in diagnosing a patient (patient classification, storing a classified patient, adapting rules, etc.). Whereas the system always found a similarity higher than 0%, he described it as an advantage because this database contains certain similarities between new and historical patients. He was also satisfied with the situation when the system found 100% similarity between new and historical patients. In general, he stated that this system could help with daily diagnostics.

The following figure 1 shows an excerpt of the created system, which offers the function of write-in a new patient, finding the coverage of a new patient and historical patients by the rules and sub-rules, and calculating similarities between new and historical patients. We can see that there is a similarity between the new patient and the historical patients; the highest achieved similarity is 60 percent.

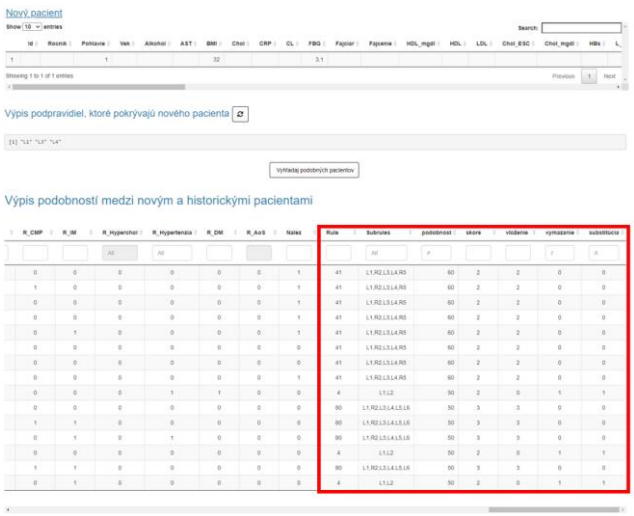


Fig. 1 Excerpt from the system in Slovak language

III. Conclusion

This paper summarized the most important parts of the dissertation. We designed a semi-automatic adaptation mechanism in the CBR process as a part of the decision support system. In the design phase, we used all our knowledge and experience from previous works of our research group. For example, we participated in decision support system creation with implemented CBR method focused mainly on the RETRIEVE phase [8]. It enables the user to find the most similar historical cases to a new case and suggest the most probable result. The system also provides useful visualizations to the expert responsible for the final decision of the new case. However, the system does not offer the possibility to adapt the rules. Also, we participated in the experiments focusing on the cluster analysis of the medical data sample from Croatia. The aim was to analyze multimorbidity from the perspective of physical frailty, cognitive impairment, and symptoms of anxiety and depression [9]. The results were published in the peer-reviewed journal Medical Science Monitor, Q3 rank by JCR. We showed that the formation of individual patient clusters is age-dependent and is influenced by the presence of mental disorders and specific cognitive performance tasks.

Acknowledgment

The work was supported by The Slovak Research and Development Agency under grants no. APVV-16-0213 and APVV-17-0550.

References

[1] M.D´Aquin, J.Lieber, A.Napoli, "Adaptation knowledge acquisition: A case study for case-based decision support in oncology," *Computational Intelligence* 2006, vol.22, pp. 161-176.

[2] N.Choudhury, "A Survey on Case-based Reasoning in Medicine," *International Journal of Advanced Computer Science and Applications* (IJACSA) 2016, vol,7 (8). doi: 10.14569/IJACSA.2016.070820

[3] P.G.W.Keen, "Decision support systems: a research perspective," *Decision support systems: issues and challenges* 1980. Oxford; New York, Pergamon Press.

[4] M.Rashidi, M.Ghodrat, B.Samali, M.Mohammadi, "Decision support systems," *Management of Information Systems* 2018.

[5] C.F.Tan, L.S.Wahidin, S.N.Khalil, N.Tamaldin *et.al.,* "The Application of Expert System : a Review of Research and Applications," *ARPN Journal of engineering and applied sciences* 2016, vol.11 (4), pp. 2448–2453.

[6] Ľ.Pusztová, F.Babič, J.Paralič. Z.Paraličová, "How to Improve the Adaptation Phase of the CBR in the Medical Domain," *International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE)* 2019, vol.3,pp. 168-177.doi: 10.1007/978-3-030-29726-8_11

[7] Ľ.Pusztová, F.Babič, J.Paralič, "Semi-Automatic Adaptation of diagnostic rules in the Case-Based Reasoning process," *Applied Sciences* 2021, vol.11 (1). doi:10.3390/app11010292

[8] Z.Tocimáková, Ľ.Pusztová, J.Paralič, D.Pella, "Case-Based Reasoning for Support of the Diagnostics of Cardiovascular Diseases," *Studies in Health Technology and Informatics* 2020, vol.270, pp. 537-541. doi: 10.3233/SHTI200218

[9] L.T.Majnaric, S.Bekic, F.Babič, Ľ.Pusztová, J.Paralič, "Cluster analysis of the associations among physical frailty, cognitive impairment and mental disorders," Intern*ational journal of Experimental and Clinical Research* 2020, vol.26.doi: 10.12659/MSM.924281

Computer Data Flow Characterization Based on Machine Learning to Streamline Network Traffic

[1]*Martin HAVRILLA (1st year)*
Supervisor: [2]*Martin CHOVANEC*

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

[1]martin.havrilla@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*— The work deals with the analysis of NetFlow records from network traffic. The growing trend of using network and other electronic communication services creates the need and emphasis on ensuring the reliability of these services. In the presented proposal, the work uses modern methods of NetFlow analysis and interprets them into a visual form. It leverages knowledge from non-relational databases and methods to analyze and process NetFlow records. This data is combined with other tools that participate in creating the final results. Among non-relational databases such as ElasticSearch, Apache Kafka, Redis, Splunk and MongoDB, the work evaluates and selects the most suitable database for its requirements. It also discusses methods of data collection as well as their subsequent analysis and visualization. Finally, the work focuses on the evaluation of possible cybersecurity and functional risks.**

***Keywords*—NetFlow, NFStream, ElasticSearch, SIEM**

I. Introduction

Nowadays, there is an increasing emphasis on the trouble-free and stable operation of network traffic in various sectors of the IT industry. To comply with the criteria in ensuring the smooth operation of network communication, services or tools to monitor the network traffic and infrastructure have been used. These monitoring tools can be divided into two categories, depending on the method of approach to monitoring, namely passive monitoring tools and active monitoring tools. The main difference between the two approaches is that passive monitoring of network traffic evaluates the problematic factor and reports the incident to the network administrator, for example, using email, SMS, push-notifications, etc. The active monitoring tools trigger with the help of artificial intelligence the process to solve this problem, for example, redirect network traffic through another router in case of a failure of one of the other routers.[1].

The most common tools for monitoring real-time network traffic include the following:

- NFStream
- NetFlow
- BWmonitor
- sFlow
- IPFIX

These protocols are used to collect and monitor packet flows. One of the best known protocols is the NetFlow protocol from Cisco Systems. This protocol is implemented directly on Cisco devices since IOS firmware version 11. It allows to create access to information on IP flows in networks collected by routers and switches. These devices gather flow data and pass it on to their collectors. The obtained data provides accurate and efficient measurements for the use of possible network resources.

The measurements obtained from the NetFlow protocol are sent to the collector and forwarded for further processing over a non-relational database.[2].

The most common form for data archiving and analysis is to use non-relational databases such as:

- ElasticSearch
- Kafka
- MongoDB
- SolarWinds

The most used non-relational database is ElasticSearch, it belongs to the category of full-text search engines based on Apache Lucene. It features a RESTful interface and among its highlights are mainly scalability, speed and high availability. The data is stored without the need to create a database schema, as the database structure is generated itself based on the inserted data.

II. Analysis of Dataflow Collection

We can understand the term of data flow as a set of IP packets passing through an observed point in the network for a predefined time interval. All packets that belong to a given flow have similar properties (e.g., source or destination address) obtained from the data occurring in the packet at the observed point. Data referred to as NetFlow service data and NetFlow data are described using flow records at the observed point of network traffic[2][3].

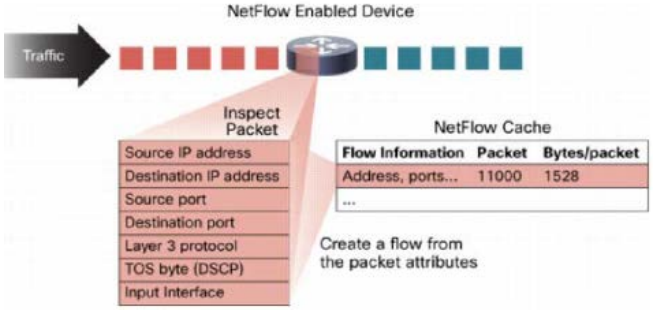


Figure 1 NetFlow flow description

Each packet that passes through the observed point, for example, router or switch is being examined. The result of the examination are attributes, these form the unique identity of the IP packet or the mark of the packet. Based on this input information is then evaluated whether it is a unique packet or a packet with a high degree of similarity to other packets under examination. All packets that have the same source and destination IP address, the same source and destination port, interface protocol, and service class are then merged into a stream, and the number of packets and the sum of bytes from this data are determined as shown in Figure 1[3].

It is then possible to extract a large set of data from the obtained NetFlow data, over which an analysis can be created. For simplicity, the following variables and their descriptions with respect to network load estimation are given[4].

- The source IP addresses present the origin of the message and they make it possible to identify the original sender of the message.
- The destination IP address represents the identity of the recipient's message.
- Source ports are involved in the characterization of applications that use network traffic.
- The service class defines the prioritization of network traffic that is affected by the QoS set in network traffic.
- The device interface defines how the traffic is used by the network device.
- The sum of packets and bytes represents the volume of network traffic.
- The subnet mask allows to specify the network prefix and define the network size.
- Flow timestamps define the life of the flow. Their use can be found when determining the number of packets and bytes per time unit.

One of the other important parts in the analysis of data collection is the Exporter. It is used to export information collected in packet flows in the network at the observed point. The collector stores this information in a predefined storage space from where the analyzer takes and then analyzes this data and generates statistics. These statistics are interpreted for the user through various tools in the form of tables, graphs, dashboards, etc.[5]. The basic topology of data collection and their subsequent analysis is shown in Figure 2.

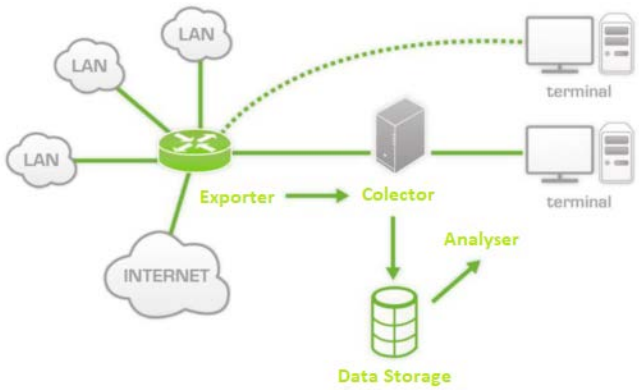


Figure 2 Block diagram of data flow collection analysis

A. Exporter

It represents a device that is used to extract data flows from network traffic. This is usually a router or switch with configured NetFlow protocol. The flow is terminated when one of the timers expires. It is either a timer that controls the passivity of the flow (inactive flow) for a predefined time period or in the case when the flow lasts longer than a defined period of activity (long flow). In general, those timers use a flow passivity setting of 15 seconds and a flow activity setting of 30 seconds. The flow is ready for export if the TCP flag indicates a flow delay[5].

NetFlow records use the User Datagram protocol (UDP) for their functionality. For proper functionality, it is necessary to configure the IP address of the NetFlow collector and the UDP port of the destination station on the source router. The use of port 2055 has become standard but in the literature it is possible to find other ports such as: 9555, 9025, 9026, 9995 and others.

To increase efficiency, routers do not store export of the flow record by default. This can cause the packet to be dropped if the network is overloaded or if the packet is damaged during transport. This problem mainly affects NetFlow v8 and also NetFlow v9 because in these versions the data flows are combined into one record and the loss of one UDP packet can have a statistical effect on some data flows[6].

At present, other transport protocols are used for transport such as SCTP to export packets and to provide sufficient resistance to packet loss. In the case of SCTP, there is a problem that interaction between each NetFlow collector and each NetFlow exporter is required which can have an adverse effect on performance limitation if the router has to work with multiple NetFlow collectors and a NetFlow collector with multiple routers.

B. Collector

This component of working with NetFlows is responsible for receiving, storing and sorting data from NetFlow records. Among other tasks which the collector covers is also pre-processing and archiving NetFlow from one or more exporters.[7].

The NetFlow collector stores these binary records in numeric format and:

- It creates compression over the data to reduce the large amount of data.
- Saves data output to databases.

C. Analyzer

The ability to characterize the network and understand its directivity (routing) is critical for ensuring availability, performance and network security. NetFlow analysis is the application of various tools to carry out the necessary monitoring, troubleshooting and due diligence, interpretation and synthesis of data flow on the network. Thorough analysis of NetFlow traffic data facilitates more accurate capacity planning and ensures that resources are used appropriately. NetFlow analysis helps to more accurately determine the point in the network where QoS needs to be set as well as to determine ways to optimize the resources used. Analyzers are also used to detect DDoS attacks and play a vital role in network security[4][9].

III. ANALYSIS OF OBTAINED DATA USING NON-RELATIONAL DATABASES

Due to the complexity of the solution of NetFlow services on which the analysis of logs will be performed it was

necessary to choose a robust solution that can collect, analyze and visualize the obtained data in real time. From the number of commercially available non-relational database solutions, it was necessary to choose a suitable database for this purpose. A choice between the following options was considered:

- ElasticSearch
- Apache Kafka
- Redis
- Splunk
- MongoDB

The main parameter for a successful selection was the processing of a large amount of data in a very short time. Among the selected databases in this parameter based on performance comparisons, ElasticSeach which currently has the most efficient search engine appears to be the best choice[5]. Another parameter for selecting a suitable database was the need to have an open source solution as in this assignment it was necessary to create a large number of filters for filtering unwanted data and also for the needs of possible implementation of other components. All databases complied with this parameter except Splunk as it is a commercial non-relational database without the open source option[4][7].

Another parameter for the selection was the possible implementation of the solution into cloud services. Among the selected databases this parameter is dominated by two non-relational databases, namely ElasticSearch and MongoDB. Both databases have, among other things, the option to connect to Amazon Web Services, Google Cloud and also Microsoft Azure.

The last parameter includes the possibility of a suitable visualization of the obtained data. All of the selected databases have data visualization or interpretation using the Grafana tool. However, due to the number of already embedded visualization tools and the wide range of community-created visualization elements, the choice fell on the ElasticSearch non-relational database[8].

Based on the analysis, the ElasticSearch non-relational database in stack together with Logstash and Kibana was selected. Due to its globally known performance parameters for processing a large number of logs in real time as well as great possibilities in the field of visualization of the obtained data. The collection system consisted of the parts shown in the block diagram on Figure 3.

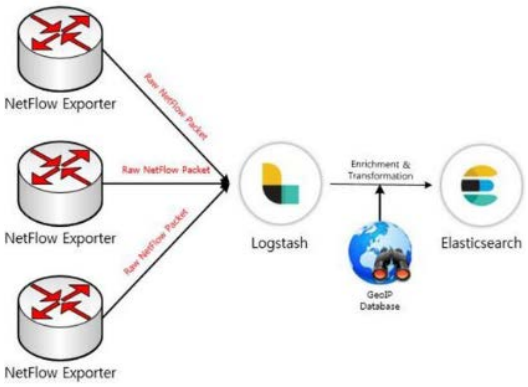


Figure 3 Connection of the Logstash collector together with the ElasticSearch analyzer

The ElasticSearch non-relational database in stack with Logstash was used for processing and analysis of the obtained data. As one of the conditions was the processing of data with high accuracy in real time, this platform provided a good basis for the implementation of the experiment. Kibana provided a suitable environment for human-friendly interpretation of data and also for their practical visualization. A block diagram of data processing, analysis and visualization is shown in Figure 4.

Figure 4 Block diagram of data processing and analysis

Analysis of data from the NetFlow protocol was provided by the Logstash collector. NetFlow collection was realized with the help of the internal Logstash module using the UDP protocol on port 2055 and subsequently interpreted into the JSON language which is used by the ElasticSearch platform.

The physical infrastructure that provides computing power consisted of two servers running Debian 10 OS and physical parameters such as 4 x CPU, 4GB RAM and 50GB HDD, and the server running the Logstash system consisted of 8 x CPU, 16GB RAM and 200GB HDD.

The obtained data can be interpreted in a visual form using the KIBANA interface. This interface provides a representation of the measured data in graphical form.

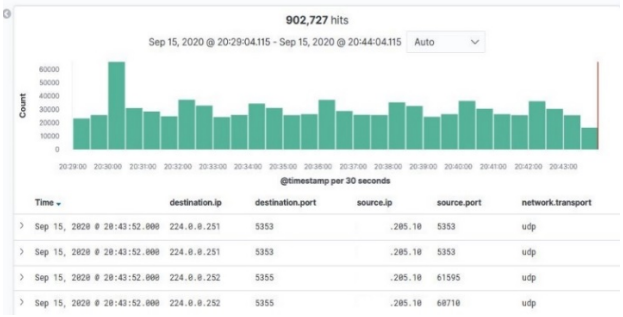


Figure 5 NetFlow in the Kibana environment

During the traffic analysis, the opportunity of recording the entire network flow was used as well as the method of recording only the network flow between devices in the environment. Using these methods there were different sizes of the index created by the ElasticSearch server. The analysis of network traffic took place in the server for 7 days. During this time, 48,274,757 network traffic records were collected. The size of these records in the database was 23.6 GB. The size and number of these records depends on the current network traffic.

The second measurement was created from a weekly record of network traffic taking place internally in the server. 23,524,228 records were collected for this type. The record reached a size of 11.4 GB.

IV. Conclusion and Future Work

The work focuses on the evaluation of theoretical knowledge in the field of working with data flows. The first part deals with the analysis of a suitable protocol for data collection. From a number of even commercially available solutions, the work selects the NetFlow protocol for its wide usability, reliability and its direct implementation in Cisco Systems devices[10]. The work also deals with the theoretical analysis of three basic components in working with data flows

and evaluates them as follows:

- The exporter is used to collect data flows from network devices. Usually the router or switch takes over the functionality of the exporter. The condition is that the NetFlow protocol is activated on such a device.
- The collector is used to store and retrieve data from the exporter. It also has functionality in sorting and pre-processing NetFlow records from one or more Exporters.

In the next part of the work, a suitable collector was selected for its purposes. The parameters that need to be met include robustness, the ability to real-time collection, processing, analysis and visualization of measured data. For long-term good results in the field of non-relational databases, the ElasticSearch non-relational database in stack with Logstash and Kibana was chosen for this task[11].

Logstash covers the functionality of the collector, the download of measured data from the exporter and further represents this data in the ElasticSearch non-relational database. The ElasticSearch non-relational database is intended to store the measured data as well as their initial analysis and to create a suitable basis for future analysis in the field of artificial intelligence. The Kibana interface covers a robust solution for visualizing the acquired data and representing them to the operator in real time. It was chosen for its wide range of data visualization options in various forms such as tables, graphs, dashboards, etc.[12].

The work that needs to be done in the next period can be divided into two parts, analysis and implementation.

The analysis part should be implemented in the field of analysis of measured data and their subsequent implementation in the field of artificial intelligence. It will use its basics in the non-relational database ElasticSearch where the SIEM interface is subsequently implemented to support the analysis of data with the help of machine learning.

Figure 6 Basic components of Elastic Stack

The resulting product of machine learning in the SIEM environment is a set of recommendations for the analyzed network flow or network device. It will be possible to re-evaluate these recommendations and, if necessary, deploy them by network administrators or make autonomous changes to networks such as QoS change, etc.

In order to implement this work, it is necessary to manage the issue of processing a large amount of data. To process such a large amount of data, it is needed to set the appropriate number of shares for the respective ElasticSearch server and also to create a connection of several servers to provide an even distribution of computing power.

Other goals in this area include the development of scripts to autonomously change the configuration of selected network functionalities such as change of QoS in the network or change of routing in case of loss or deterioration in data transmission quality.

The last goal is to create a clear and functional environment for network administrators in which it will be possible to observe and also implement changes.

ACKNOWLEDGMENT

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF.

REFERENCES

[1] BILGE, Leyla, et al. Disclosure: detecting botnet command and control servers through large-scale netflow analysis. In: Proceedings of the 28th Annual Computer Security Applications Conference. 2012. p. 129-138.

[2] HOFSTEDE, Rick, et al. Flow monitoring explained: From packet capture to data analysis with netflow and ipfix. IEEE Communications Surveys & Tutorials, 2014, 16.4: 2037-2064.

[3] CARELA-ESPAÑOL, Valentín, et al. Analysis of the impact of sampling on NetFlow traffic classification. Computer Networks, 2011, 55.5: 1083-1099.

[4] BIN, Liu, et al. A NetFlow based flow analysis and monitoring system in enterprise networks. Computer networks, 2008, 52.5: 1074-1092.

[5] ESTAN, Cristian, et al. Building a better NetFlow. ACM SIGCOMM Computer Communication Review, 2004, 34.4: 245-256.

[6] ČERMÁK, Milan, et al. A performance benchmark for NetFlow data analysis on distributed stream processing systems. In: NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2016. p. 919-924.

[7] SANTOS, Omar. Network Security with NetFlow and IPFIX: Big Data Analytics for Information Security. Cisco Press, 2015.

[8] WANG, Yuan-Ting, et al. The Implementation of NetFlow Log System Using Ceph and ELK Stack. In: International Conference on Frontier Computing. Springer, Singapore, 2018. p. 256-265.

[9] YANG, Chao-Tung, et al. On construction of a network log management system using ELK Stack with Ceph. The Journal of Supercomputing, 2019, 1-17.

[10] TIAN, Zehua. Management of large scale NetFlow data by distributed systems. 2016. Master's Thesis. NTNU.

[11] LAHMADI, Abdelkader; BECK, Frédéric. Powering monitoring analytics with elk stack. In: 9th International Conference on Autonomous Infrastructure, Management and Security (AIMS 2015). 2015.

[12] ALMEIDA, E. S.; KOGA, I. K. A framework to analyze massive data from applications and services of a meteorological data center. In: IV world congress on systems engineering and information (WCSEIT), Science and Education Research Council (COPEC), Guimarães, November. 2017. p. 26-29.

Analysis and Evaluation of Data Using Artificial Intelligence for Cybersecurity

[1]*Martin HASIN (1st year),*
Supervisor: [2]*Martin Chovanec*

[1,2]Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic

[1]martin.hasin@tuke.sk, [2]martin.chovanec@tuke.sk

***Abstract*—The article focuses on the research and design of the infrastructure used for data collection from the network traffic. The aim was to identify the structure of network traffic that occurs in the measured network. From the mentioned data it is possible to evaluate network security of the protocols used in communication and subsequently compare the obtained data with the security standards introduced in the experimental platform. The measured data was then stored in the non-relational database Elasticsearch. The selected database is intended for the analysis of a large amount of data. The most important part is the creation of an interactive web environment in order to interpret the measured data in a user-friendly form. The next step in the processing of measured data involves the application of artificial intelligence methods to evaluate and predict the existence of potential security problems.**

***Keywords*—Nfstream, Elasticsearch, NetFlow, network monitoring**

I. Introduction

Network monitoring is a concept that is becoming necessary in rapidly evolving data networks. It is possible to monitor data infrastructure actively or passively. Active monitoring is most often associated with systems for monitoring network elements and subsequent reporting of faults in real time via SMS or EMAIL messages. This kind of monitoring does not allow administrators to have sufficient knowledge of what is happening on a given network in real time [1]. The monitoring system itself usually consists of several parts that monitor the network infrastructure at different levels. At least one of these devices has the ability to store measured data to provide deeper analysis in the network. Based on these requirements, various applications were created to monitor network traffic in real time. There are several monitoring options available: NetFlow, ntopng, BWmonitor, NFStream...

II. NetFlow protocol

The most used protocol is NetFlow [2]. This protocol developed by CISCO was to be an additional service on their routers. The main use of this protocol is to monitor network traffic based on IP network flows through which it is possible to get a real-time view of the network. During development, various versions of the NetFlow protocol were created [3]. The first version released in 1996 provided only basic network data. These data included IP address information only in the IPv4 protocol. The development of this protocol has progressed and the current version 10 has received support for various protocols such as IPv6, MPLS, BGP. This latest version is called IPFIX [4].

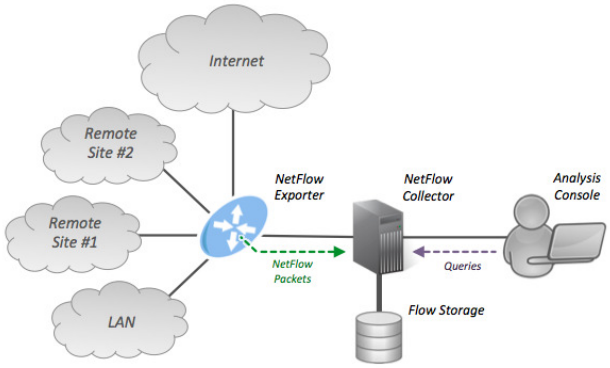


Fig. 1. Typical connection of NetFlow collector

A typical NetFlow connection includes the following components:

- Flow exporter
- Flow collector
- Analysis console

As the network infrastructure developed rapidly, the need to monitor the network using NetFlow increased. This also increases NetFlow collector speed requirements as well as capacity requirements for data storage and analysis. At high speed data infrastructure, there is a need to monitor the network using NetFlow as an external service where it is possible to fully use the speed of the network card and the speed of the processor that is needed to calculate NetFlow [5]. It is also necessary to store the measured data for their amount in non-relational databases which are faster in these deployments than relational databases [6].

III. Architecture of network traffic measurement using NFStream

NFStream framework is a component that uses the Python programming environment [7]. This framework allows you to analyze the data that enters the program in two ways. The first way is to use a PCAP file that was created on a given infrastructure [8]. However, this approach is an offline network analysis solution. It is suitable for Machine Learning exercises with predefined patterns of unwanted traffic in the network. The second way of analysis is to use ONLINE reading of

data from the network adapter using the libpcap library [9]. Using the Port Mirror technology, the required network flow from the production network adapter where the measured traffic takes place is mirrored to the given network adapter. Using the NFStream framework, it is possible to quickly and flexibly analyze data in real time with minimal delay [10]. The delay can be eliminated by increasing the performance of the processor on which the analysis takes place. It is necessary to have the highest performance on individual processor cores. Performance is also affected by the given network adapter which must include the distribution of network traffic to multiple processor cores. The NFStream framework allows the measured data to be processed and analyzed in real time. It is possible to use custom plugins with which it is possible to create custom attributes for NetFlow export. NFStream framework has the following functionalities:

- High speed - this framework is designed for quick network analysis using the libpcap library.
- L7 analysis - the framework enables in-depth packet inspection using nDPI. This process allows you to quickly determine the type of application through which it communicates using metadata (TLS, SSH, DHCP, HTTP ...)
- Statistical functions - the framework allows you to record statistical information about network flows which are calculated only after the connection is terminated. Most common statistics: sequence of first n packets sizes, inter arrival times and directions.
- Flexibility - the framework allows you to create your own modules with which it is possible to add new necessary functionalities in the Python language.
- Machine Learning - outputs from this framework can be implemented in Machine Learning models using Python libraries [11]. The resulting trained models can be quickly implemented into real operation using NFPlugin.

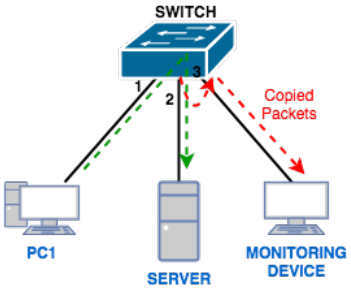


Fig. 2. Connection of NetFlow probe in infrastructure

The connection of this probe makes it possible to reduce the load on the main router by the fact that the analysis itself takes place on the server side. This analysis has the advantages of scaling and modifying the NetFlow filter itself. The NetFlow process starts separate processes that perform separate measurements.

In fig. 3 is a block diagram of the measurement process. Network traffic that replicates to a server is broken down by a model into a number of processes. With this approach, it is possible to analyze network traffic without packet loss even with higher loads. By distributing the measurement itself into processes, the power is automatically distributed among a larger number of CPU cores of a given server [12]. The measurement marked Meter 0 consists of two blocks. The first block is packet analysis. These packets are grouped by

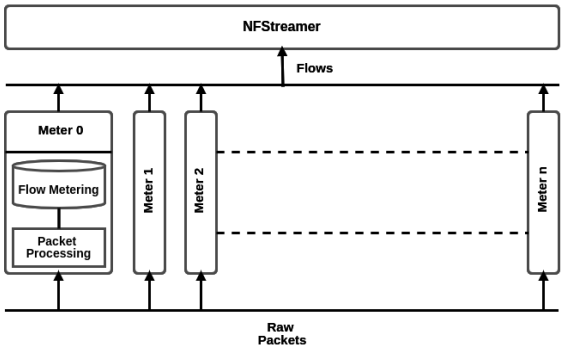


Fig. 3. NFStream topology

NetFlow into blocks of 1500 bytes and then packet-processed into the form required for NetFlow. The detected data is stored in the Cache in RAM memory which prevents disk storage delays. The results from this data are then sent to the NFStreamer application. This process allows you to control and set the created meter processes. The outputs from these processes can be further processed in Python. Using the NFStream framework technology, it is possible to divide the storage of measured data into processes. These measured data are written to a dedicated medium in several processes where the processing process is accelerated.

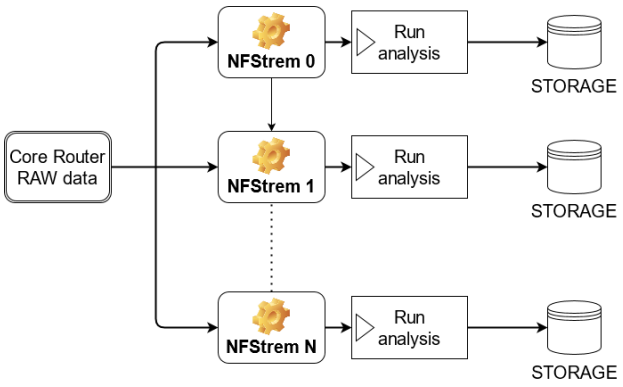


Fig. 4. NFStream data analysis topology

Fig. 4 ishows the topology by which the measured data can be analyzed and subsequently stored on a storage. The analysis supported by the NFStream framework is the use of the Packet Inspection Library nDPI which can determine from the measured network traffic what kind of application communicates in a given network. [13] This library is regularly expanded with new protocols that arise in network traffic. When used, the library allows you to turn off any part of the network traffic monitoring analysis that is not needed to evaluate it, and thus it is possible to reduce the system load. The nDPI library allows you to search for protocols that can communicate on the network even on non-standard ports since the concept where port = given application no longer applies [14]. With the increasing trend of encrypted communication, the nDPI library contains a decoder for SSL certificates. Using this decoder it is possible to detect the protocol and the application also in encrypted communication such as Apple iCLOUD, Google CLOUD, Citrix ... These protocols would otherwise be undetectable.

IV. DATA ANALYSIS USING NON-RELATIONAL DATABASES

Data analysis is the process of analyzing various attributes and summarizing them into useful information. It is about extracting useful information from large databases. From the collected data it is possible to create various models such as:

- Prediction models - models based on collected historical data can predict possible behavior in the future.
- Segmentation models - a model in which the goal is to classify the amount of data into a smaller number of groups that describe the given research area.

Data analysis is usually performed on data that is stored in the appropriate database. It is possible to choose two types of databases depending on the amount and nature of data to be analyzed:

- Relational database - data in this database are stored in tables that are interconnected. The links between the tables are used to search for the required information. A table is a set of data that focuses on a single area and consists of columns and rows.
- Non-relational database - the data in this database do not use traditional relational database models. The data in this database can be in structured or unstructured form. Searching in these databases is based on searching for terms in stored data using JSON. These databases are designed for highly optimized data reading and retrieval. Usually, these databases do not have as many functionalities as SQL databases but they have high performance in data storage and indexing.

A. Architecture of Elasticsearch Non-Relational Database

Elasticsearch non-relational database is focused on full text searches based on Apache Lucene. The database uses the Full Rest API for communication where it allows you to create high availability, speed and scalability [15]. The non-relational database is developed in the JAVA programming language which allows it to run on various platforms. Data search is possible via the KIBANA WEB interface. The main advantage of this database is the high speed of searching and applying filters.

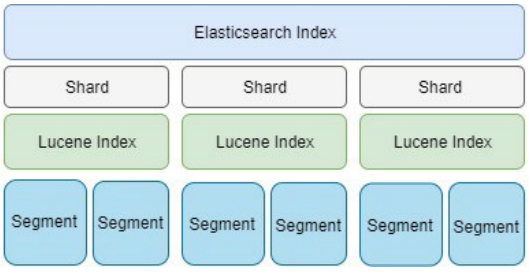


Fig. 5. Elasticsearch structure

Elasticsearch non-relational database stores all data in an index that can be compared to a database in traditional relational databases. Each index in the Elasticsearch database contains the following parameters:

- Shard - when using a large amount of data, the index is divided into several smaller fragments. This division allows load and capacity to be evenly distributed among several Elasticsearch servers [16].
- Lucene Index - this index is a reverse version of the Apache Lucene library index from the classic index. This index consists of several parts with inverted indexes.
- Segment - is a place on the disk where data from the Lucene index is written. This data is read from the cache memory and written to segments. Searching in the data that is stored in the cache memory is not possible, therefore the data is periodically copied into segments.

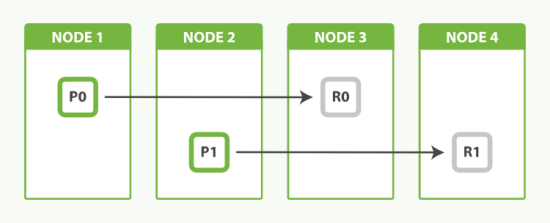


Fig. 6. Elasticsearch cluster data structure

To achieve high availability in the Elasticsearch database, the index is divided onto several nodes. Nodes are servers that are connected in a cluster solution and form one unit. The cluster itself can be formed by just one node. However, with such a cluster, there is no option to create system redundancy. For multiple nodes in a cluster, at least one master server must be selected to control the other servers. This server controls the creation or deletion of indexes on other nodes in the cluster. As with Apache Solr, the Elasticsearch database has the index distributed into several smaller shard blocks. These blocks are distributed over some nodes [17]. The cluster contains the following shards:

- Primary shards - is a group of shards whose group belongs to one index.
- Replica shards - is a replica of all primary shards. Replicated shards are divided according to indexes where it is possible to repair a damaged primary shard from unapplied data.

V. CONCLUSION AND FUTURE WORK

The beginning of the work was to find a suitable mechanism by which data would be collected from the data network. Several variants were tested where the result was the NetFlow protocol using NFStream Python libraries. By measuring the stability of speed and modularity of applications, the NFStream framework has emerged as sufficiently modular also for future development. It was also possible to incorporate the export of data into the Elasticsearch non-relational database using this framework. Writing to this database is done by the created module which communicates using JSON. The database consists of a cluster which consists of 6 servers where 2 servers are referred to as master servers. To increase the search speed, the number of shards was set to 20. This increased the search speed but also increased the number of replicas in the cluster. Measurements have shown that a larger amount of shards in the cluster increases the load on disk speed as well as network throughput.

The database cluster is built for a large number of searches in the database as the database will consist of an index containing NetFlow records and an index containing Threat samples. The data process will evaluate real data with already known threats that have occurred in the measured network. This evaluation will speed up the analysis of NetFlow records, as

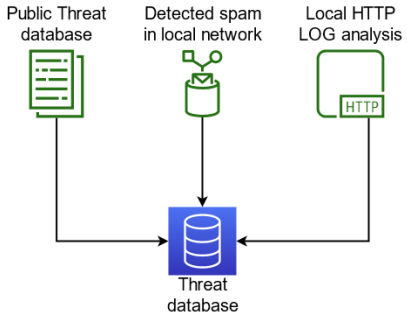


Fig. 7. Threat database structure

they will not have to analyze these records using the Machnine Learning model, they will only be automatically marked with a known network threat value. The Threat database will consist of a set of data that will be measured on application servers in a given network. The following data can be analyzed from the given network:

- Spam Filter database - analysis of messages that pass through the spam filter in the given network. These messages are evaluated in a given filter by a certain degree of threat. This database contains a large amount of data on infected networks from which SPAM is spread in mail communication.
- HTTP LOG - analysis of web services that are in the given network. This makes it possible to use access logs to detect various types of attacks or scanning robots that scan content on given servers and look for vulnerabilities.
- Public Threat database - a public database of known IP addresses from which unwanted activity occurs. Devices behind these IP addresses are being misused to create unwanted activity on other public networks.

By combining these three databases, it is possible to very quickly search for threats in a given network in the current network traffic. This data changes in the database according to the current data from these application servers. This makes it possible to quickly respond to and limit various network scans or prevent a DDoS attack on a given service.

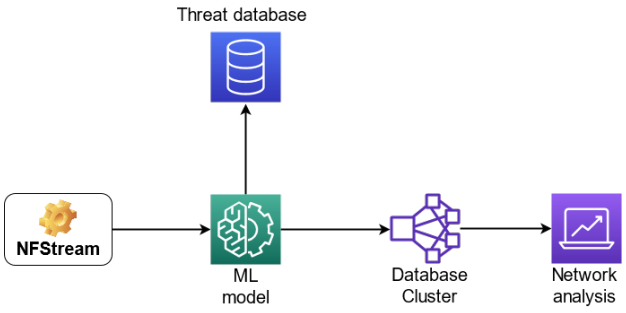


Fig. 8. NetFlow analyzer

Fig. 8 shows a network analysis model by which measured NetFlow can be evaluated in real time. The main point is to create a Machine Learning model that would automatically evaluate and predict on untrained samples of network traffic how dangerous the measured communication is for a given network. The input into the decision-making of artificial intelligence is entered by the Threat database which contains samples of known threats in the network, as well as globally known threats. These data help to create a better ML model that evaluates the network. The result is a network analysis that evaluates the detected threats and then uses them to intervene in the given network. Using data from this database, it is possible to create rules on network elements and then capture and analyze the suspicious network traffic more closely. This makes it possible to achieve a minimal response between the beginning of the attack and the very restriction of the communication by means of security elements in the network. Another type of output is a graphical interpretation of data where each measured NetFlow will be represented by the value of security risk for a given network. Using this data, it is then possible to create a clear report in graph form of the network state at any time.

Acknowledgment

This publication has been published with the support of the Operational Program Integrated Infrastructure within project: Research in the SANET Network and Possibilities of Its Further Use and Development (ITMS code: 313011W988), co-financed by the ERDF.

References

[1] A. Pekár, M. Chovanec, L. Vokorokos, E. Chovancová, P. Feciľak, and M. Michalko, "Adaptive aggregation of flow records," *Computing and Informatics*, vol. 37, no. 1, pp. 142–164, 2018.

[2] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of data center traffic: measurements & analysis," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, 2009, pp. 202–208.

[3] C. Estan, K. Keys, D. Moore, and G. Varghese, "Building a better netflow," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 245–256, 2004.

[4] B. Claise, M. Fullmer, P. Calato, and R. Penno, "Ipfix protocol specification," *Internet-draft, work in progress*, 2005.

[5] M. Kiran and A. Chhabra, "Understanding flows in high-speed scientific networks: A netflow data study," *Future Generation Computer Systems*, vol. 94, pp. 72–79, 2019.

[6] C. Gyorödi, R. Gyorödi, and R. Sotoc, "A comparative study of relational and non-relational database models in a web-based application," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 11, pp. 78–83, 2015.

[7] L. Deri and N. SpA, "nprobe: an open source netflow probe for gigabit networks," in *TERENA Networking Conference*, 2003, pp. 1–4.

[8] M. J. Sweeney, "A framework for scoring and tagging netflow data," 2019.

[9] M. P. Karpowicz and P. Arabas, "Preliminary results on the linux libpcap model identification," in *2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE, 2015, pp. 1056–1061.

[10] N. Shazeer, Y. Cheng, N. Parmar, D. Tran, A. Vaswani, P. Koanantakool, P. Hawkins, H. Lee, M. Hong, C. Young *et al.*, "Mesh-tensorflow: Deep learning for supercomputers," *arXiv preprint arXiv:1811.02084*, 2018.

[11] J. Hou, P. Fu, Z. Cao, and A. Xu, "Machine learning based ddos detection through netflow analysis," in *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*. IEEE, 2018, pp. 1–6.

[12] C. N. X. TONGGE, "Study on netflow-based ip traffic aggregation in high-speed network [j]," *Microcomputer Information*, vol. 3, 2008.

[13] J. Sherry, C. Lan, R. A. Popa, and S. Ratnasamy, "Blindbox: Deep packet inspection over encrypted traffic," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 2015, pp. 213–226.

[14] G. B. Satrya, F. E. Nugroho, and T. Brotoharsono, "Improving network security-a comparison between ndpi and l7-filter," *International Journal on Information and Communication Technology (IJoICT)*, vol. 2, no. 2, pp. 11–11, 2016.

[15] C. Gormley and Z. Tong, *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. " O'Reilly Media, Inc.", 2015.

[16] R. Kuc and M. Rogozinski, *Elasticsearch server*. Packt Publishing Ltd, 2013.

[17] P. M. Dhulavvagol, V. H. Bhajantri, and S. Totad, "Performance analysis of distributed processing system using shard selection techniques on elasticsearch," *Procedia Computer Science*, vol. 167, pp. 1626–1635, 2020.

Deep neural networks for speech-to-text systems

[1]*Slavomír GEREG (2nd year)*

Supervisor: [2]Jozef JUHÁR

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]slavomir.gereg@tuke.sk, [2]jozef.juhar@tuke.sk

***Abstract*—The main topic of this article is automatic speech recognition system Kaldi and my recent work with it. At the beginning is shortly described automatic speech recognition usage in the modern user applications. Deep neural networks and automatic speech recognition system Kaldi are briefly introduced after that. The main part of this article is my work with automatic speech recognition system Kaldi. The description of my previous research in these field, my current state of research and my future goals are also presented there. In conclusion there are briefly outlined topics I would like to deal with in my dissertation.**

***Keywords*—automatic speech recognition, deep neural networks, Kaldi toolkit**

I. INTRODUCTION

Automatic speech recognition or speech to text transcription is the complex process of converting of audio speech signal into the it's corresponding text form. Systems and algorithms for automatic speech recognition have been important part of research for more than 6 decades. Currently, there are many commercial products and applications that are using automatic speech recognition systems and algorithms. One of the most common uses of these algorithms and systems for the general users are applied in voice control or automatic systems. In the past automatic speech recognition has been used in a variety of applications for example text dictation, gaming industry, voice dialing, voice search, interactive voice dialogue, call routing, voice control, official document transcription and robotics. Today we can implement and use these automatic speech recognition systems and algorithms on different types of devices, computers, servers or electronics. As computer technology grows and improves, even more sophisticated automatic speech recognition services and applications are becoming a reality. For example we can use the extensive use of voice assistants in smart mobile phones (Google assistant, Siri, Cortana, Alexa, ...), smart home voice control applications and products (voice assistants from Google, Amazon, ...), voice control in home gaming systems (Kinect), voice navigation and many other advanced applications and systems that use automatic speech recognition [1], [2], [3], [4], [5], [6], [7].

Popular approach in modern speech recognition system these days is usage-based on DNN. Deep neural networks are composed of neurons that are interconnected. The neurons are arranged in layers. The first layer is the input layer that corresponds to the data properties. The last layer is the output layer, which provides probabilities of classes or markings (classification task). The output of the y-th neuron is calculated as the non-linear weighted sum of its input [9].

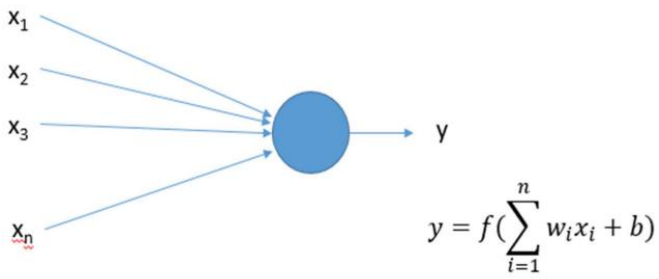


Figure 1 Example of one neuron and its connection [9]

II. MY PREVIOUS RESEARCH

In my previous work on diploma thesis I was making a test database for automatic speech recognition system Kaldi. In this database I was using audio recording from meetings and sessions. These data had a lot of background noise and audio interference included into recordings. The creation of this test database consisted a lot of, from manual annotation of acoustic data and records, and modification of these annotated data into a form suitable for testing by the automatic speech recognition system Kaldi. I also experimentally tested impact of the training data on the result accuracy of automatic speech recognition provided by automatic speech recognition system Kaldi. I tried to compare results of automatic speech recognition process with training from different train data. I was comparing training with data from LRMT KEMT TUKE laboratory database and after that I have tried to extend this database with high amount of data (247 hours 31 minutes and 57 seconds) automatically recognized by SARRA (speech transcription system created in LRMT KEMT TUKE)[1]. My hypothesis was based on the assumption that a higher volume of automatically recognized training data of the same type as test data (meeting audio, the same acoustic channel, same recording environment, ...) should result in a lower error rate in automatic speech recognition system Kaldi. After that in

[1] https://marhula.fei.tuke.sk/sarra/login

first year of my Phd. study I was trying study a theory of automatic speech recognition and some modern approaches based on using neural networks in automatic speech recognition systems. These approaches are used in automatic speech recognition system Kaldi too.

III. CURRENT STATE OF RESEARCH

In my second year of mu Phd. study I have started some practical tests and experiments with automatic speech recognition system Kaldi on my own. In my previous work on my diploma thesis I have tried to prepare data for automatic speech recognition system Kaldi, and I evaluate its results, but I have never tried to work with automatic speech recognition system Kaldi on my own. So, in my work I start with studying practical use of this system. I went through various stages of work with automatic speech recognition system Kaldi for example preparing of acoustic data to suitable form, finding best acoustic model for training, system training with real database, testing and evaluation of results provided by system Kaldi. I also created a manual about practical system Kaldi training for my colleagues and for future research.

IV. SPEECH RECOGNITION TOOLKIT KALDI

Kaldi is an open-source toolkit for automatic speech recognition. It is written in C ++ and licensed under the Apache License v2.0. Kaldi's goal is modern, flexible and easily extensible code that is easy to understand and modify. Kaldi is available for download at SourceForge[2]. Kaldi can be compiled on commonly used Linux-based systems as well as on Microsoft Windows-based systems [10], [11], [12].

The Kaldi toolkit allows to use many specific requirements for automatic speech recognition systems, such as structure based on Finite-State Transducer (FST), extensive linear algebra support and non-limiting license [10], [13].

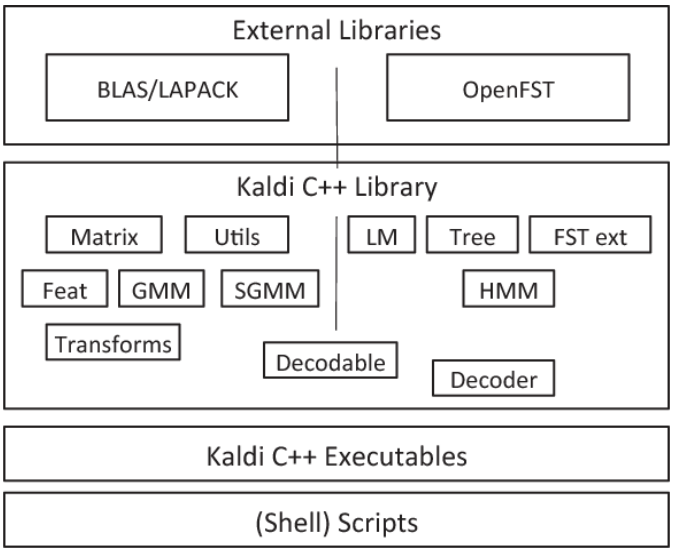


Figure 2 Basic block diagram for Kaldi automatic speech recognition [10]

In Figure 2 we can see the basic block diagram of the automatic speech recognition system Kaldi.

Library modules can be grouped into two different groups, each of these groups depends only on one of the external libraries. These two groups are bridged by the DecodableInterface module [10].

The modules shown in Figure 2 below are always dependent on one or more modules above them.

[2] Available at: http://kaldi.sf.net/

V.FURTHER RESEARCH DIRECTION

In my future work I want to continue with my previous experimental work, and I want to experimentally evaluate effect of training database to automatic speech recognition system Kaldi effectiveness. I want to find a right amount of automatically transcribed data in training database of automatic speech recognition system Kaldi and develop a methodology for speech recognition training with this kind of database. In this step I want to focus on the success of speech recognition but also the computational and time-consuming system training.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project VEGA 1/0753/20, KEGA 009TUKE-4/2019 and by Slovak Research and Development Agency project APVV-15-0517.

REFERENCES

[1] Deng L., O'Shaughnessy D., "Speech Processing-A Dynamic and Optimization Oriented Approach," Marcel Dekker Inc., New York, 2003.

[2] Huang X., Acero A., Hon H.W., "Spoken Language Processing," Prentice-Hall, Upper Saddle River, NJ, 2001b.

[3] Baker J., Deng L., Glass J., Khudanpur S., Lee C.H, Morgan N. et al., "Research developments and directions in speech recognition and understanding," Part I. IEEE Signal Process. Mag. 26 (3), 2009a ,75-80.

[4] Davis K.H., Biddulph R, Balashek S., "Automatic recognition of spoken digits," J. Acoust. Soc. Am. 24 (6), 1952, 627-642.

[5] He X., Deng L., "Speech-centric information processing: An optimization-oriented approach," Proc. IEEE 101 (5), 2013, pp. 1116-1135.

[6] Li J., Deng L., Haeb-Umbach R., Gong Y., "Robust automatic speech recognition: A bridge to practical applications," 2015. ISBN 978-0-12-802398-3.

[7] Juhár J. et al., "Rečové technológie v telekomunikačných a informačných systémoch." Košice: Equilibria, s.r.o., 2011. ISBN 978-80-89284-75-7.

[8] Bundzel M., Sincak P., "Combining gradient and evolutionary approaches to the artificial neural networks training according to principles of Support Vector Machines." In: Proceedings of the IEEE International Joint Conference on Neural Network, Vancouver, Canada, July 16-21, 2006, DOI: 10.1109/IJCNN.2006.246976.

[9] Fohr D., Mella O., Illina I., "New Paradigm in Speech Recognition: Deep Neural Networks." IEEE International Conference on Information Systems and Economic Intelligence, Apr 2017, Marrakech, Morocco. 2017.

[10] Povey D., Ghoshal A. et al., "The Kaldi Speech Recognition Toolkit." In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US, 2011. ISBN 978-1-4673-0366-8

[11] Smit P., Virpioja S., Kurimo M., "Improved Subword Modeling for WFST-Based Speech Recognition." In: Annual Conference of the International Speech Communication Association (INTERSPEECH), Stockholm, August 2017, pp 2551–2555.

[12] Alyousefi S. H., "Digital Automatic Speech Recognition using Kaldi." MSc. Thesis, Florida Institute of Technology, Melbourne, Florida, May 2018.

[13] Enarvi S., Smit P., Virpioja S., Kurimo M., "Automatic Speech Recognition with Very Large Conversational Finnish and Estonian Vocabularies." IEEE/ACM Transactions on Audio, Speech, and Language Processing, volume 25, issue 11, November 2017, pages 2085–2097.

Influence of Multiple Reflows on Electrical Resistivity of Solder Joints

[1]*Daniel DZIVY (2nd year)*
Supervisor: [2]Alena PIETRIKOVA

[1,2]Dept. of Technologies in Electronics, FEI TU of Kosice, Slovak Republic

[1]daniel.dzivy@tuke.sk, [2]alena.pietrikova@tuke.sk

***Abstract*—This paper is focused on measuring of the electrical resistivity of the solder joints. It also compares the influence of the most popular surface finishes and multiple reflows on quality of the solder joints. This paper presents the possibility of evaluation the quality of the solder joints based on SAC305 by measuring of its electrical resistance. For the purpose of the electrical resistivity measurement, the design for measuring of very small resistance of solder joints was developed.**

***Keywords*—Surface finishes, solder joint, electrical resistivity, multiple reflows, soldering.**

I. Introduction

Nowadays, the low-cost applications, the miniaturization of the electrical components, the ability of a mass production and environmentally friendly materials are the main reasons for the improving a manufacturing process. One of the most important steps in a manufacturing process is the right choice of the surface finish of the PCB, which interacts with the solder and it is affecting for electrical properties and microstructure of the solder joint. During the manufacturing process of the PCB, the complexity of the design sometimes requires multiple reflow processes. It is very important what is the quality of surface finish on PCB that influences the development of microstructure and solder joint quality. The electrical resistivity of the solder joint is one of the most popular parameters that is usually analyzed in this connection. Electrical resistivity is changing after every reflow and it is very important what is the stability of the surface finish on PCB. In the manufacturing, the maximal amount of multiple reflows is various, but generally it will not reach over 6 [1], [2], [3].

For this purpose, we made PCB design accommodated for measuring the small changes of electrical resistance of the solder joints.

II. Initial Status

During the second year of my PhD study, I was analyzing the surface finishes of the PCB. Also, I advised with the basic manufacturing processes of surface finishes of the PCB, I analyzed the possibilities of analyzing the quality of PCB surface finishes and I analyzed the possibilities of improving the properties of PCB surface treatments for the purposes of use in electronics. The main accent in my work was devoted to theoretical analysis of the solder fluxes influence on the surface free energy in dependence on the surface finish of the PCB and the type of the solder flux. The analysis of the plasma influence on the surface free energy in dependence on the surface finish of the PCB and the plasma was realized as well. We were also studying about analysis of an adhesion of the solder joints from the perspective of the surface finish and wetting. Theoretical research about influence of the thickness of a reactive zone on border line between the surface finish of the PCB and the solder on the mechanical and electrical properties were also realized. The aim of our work was optimalization of customization of the PCB surface from the perspective of the solder joint quality and long-term reliability.

III. Solved Tasks in The Previous Year

Tasks which are summarized in the following section were solved in the last year of postgraduate study.

A. *Development of the PCB design optimized for measuring very low electrical resistances*

The aim of this task was to develop the new PCB design optimized for measuring the very low electrical resistances. The design for the testing samples using of the Kelvin´s 4-wire method was developed and used for all types of the surface finishes. The conductive pattern (Fig. 1) was applied in fabrication process of PCBs with different surface finishes (HASL Sn, HASL Pb, ENIG, SnAg, pure Cu). For all samples, surfaces were purified in the same manner before applying the surface finishes (standard methods). Design of the circuit had to be optimized for the accurate and fast measuring of the electrical resistance. Test pattern of the bottom and also top parts are in Fig. 1. One test board consists of 21 samples and 42 solder joints, where the electrical resistance can be measured. One test sample consists of 4 solder joints, which were formed by soldering the top part on the bottom PCB as shown in the Fig. 1. For all top part´s surface finish was selected ENIG because it is stable due to multiple reflows as well as it is reliable. Dimension of one solder pad is 1.6 mm x 0.9 mm. Thickness of the solder joint is various because every surface finish has different thickness.

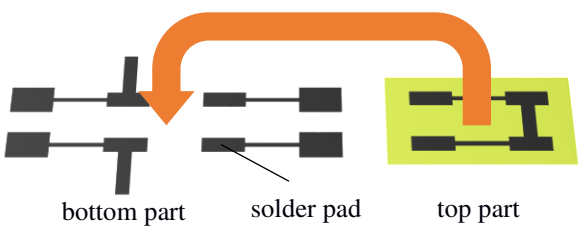


Fig. 1 Test pattern of the bottom part of PCB together with 4 point top part

Measuring of electrical resistivity was realized by measuring of the electrical resistance with Agilent 4338B miliohmmeter and counting values of the electrical resistivity ρ by the formula:

$$\rho=R.S/l \ [\Omega.m], \quad (1)$$

where: R – measured electrical resistance, S – area of copper pad, l – thickness of copper pads together with various surface finishes and solder.

Figure 2 shows the equivalent resistor values for the solder joint. Fig. 2 also shows measurement´s principle of the solder joint. For the measurement, current probes were connected with 10 mA through two pins on the opposite sides of the PCB. Voltage was measured using a detached two-pin probe, that was relocated after every measured value. Measured resistance is combination of 5 layers as shown in Fig. 2:

$$R=R_C+R_{TSF}+R_S+R_{BSF}+R_C, \quad (2)$$

where: R_C – copper´s layer resistance, R_{BSF} – surface finish´s layer resistance of the bottom part, R_S – solder´s layer resistance and R_{TSF} – surface finish´s layer resistance of the top part.

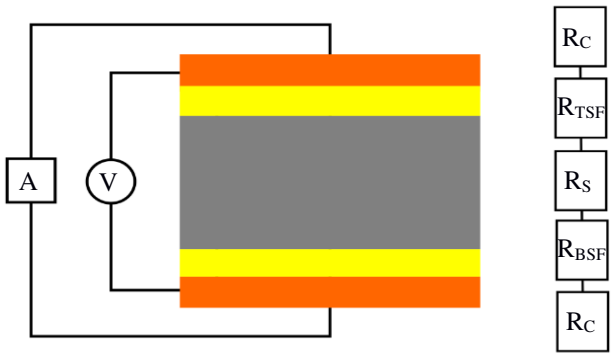


Fig. 2 Principle of the measurement of the solder joint´s electrical resistance a) and it´s alternative resistor model b)

For soldering, solder paste based on SAC 305 with no clean solder flux was chosen, because it is the most popular composition of the lead-free solders. The thickness of the solder depends on the thickness of the stencil, which is 127µm. For top parts, as was mentioned, the ENIG surface finish was chosen. The only variable is resistance of the bottom parts surface finish. This is due to the different electrical resistivity of the metals used for the production of surface finishes. Also, the use of different metals affects the microstructure, which results in a change in electrical resistance [4], [5], [6].

B. Influence of multiple reflows on the solder joint electrical resistivity

The aim of this task was to prove, that the multiple reflows from 1 to 6 times influence the electrical resistivity of the solder joints. We applied IR furnace with reflow profile that is recommended by solder paste producer.

In the Fig. 3, the results show, that the electrical resistivity of the solder joints is changing with the type of the surface finishes. The electrical resistivity of the surface finish based on SnAg is the lowest. For the ENIG and the Cu surface finish, the variance of the electrical resistivity is minimal. After second reflow, electrical resistivity decreases. After third reflow, electrical resistivity increases much higher. Even after 5 reflows, the electrical resistivity of the SnAg is low and for ENIG and Cu is not so high. The HASL surface finishes have the biggest values of the electrical resistivity. The HASL Pb has higher value due to the higher electrical resistivity of the lead.

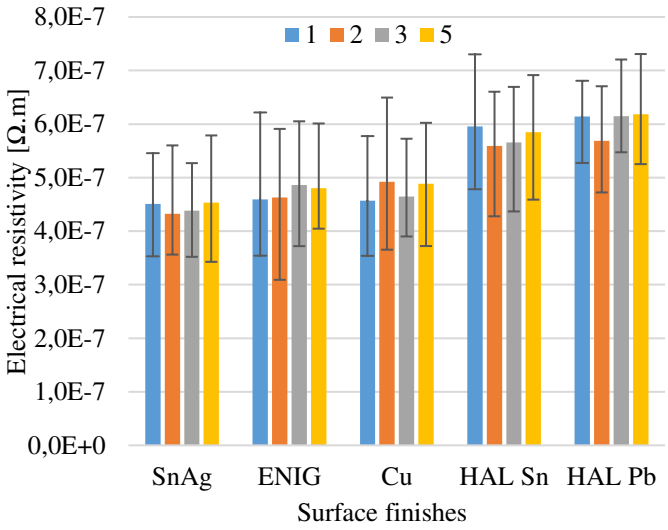


Fig. 3 Graphical representation of the average values of electrical resistivities of the solder joints after 1, 2, 3 and 5 reflows

IV. Future Work

In the future work we will focus on the surface treatment using the various types of the solder fluxes, surface treatment using a cold atmospheric air plasma, analysis of influence of the multiple reflows on the shear test strength in correlation with the electrical properties and height of the IMCs layer.

V. Conclusion

This research was focused on influence of multiple reflows on the solder joints electrical resistivity and influence of the various types of surface finishes on the electrical resistivity. This research shows, that the surface finish based on SnAg has the lowest value of electrical resistivity of the solder joint even after multiple reflows. Our measurements show that this surface finish has the most stable properties after thermal exposure. For HASL surface finishes, the electrical resistivities are much higher in compare to other surface finishes.

References

[1] TRESCASES O.: Guide to Designing and Fabricating Printed Circuit Boards, University of Toronto, Jan. 2006.

[2] MIN. Z. – QIU Y.: Interfacial reaction and IMCs formation between Sn-0.7Cu solder and Cu substrate during reflow soldering, 16th International Conference on Electronic Packaging Technology (ICEPT), Changsha, China, Aug. 2015.

[3] PANDHER R. – PACHAMUTHU A.: Effect of multiple reflow cycles on solder joint formation and reliability, 2010.

[4] NOH B.I.: Effect of number of reflows on the mechanical and electrical properties of BGA package, *Intermetallics*, vol. 14, pp. 1375-1378, Oct. 2006.

[5] LIU D. S. – NI C. Y.: A study on the electrical resistance of solder joint interconnections, *Microelectronic Engineering*, vol. 63, pp. 363-372, 2002.

[6] HIRMAN M.: Electrical resistance of Solder Joints on Conductive Ribbons, *43rd International Spring Seminar on Electronics Technology (ISSE)*, Košice,

Research on the impact of electric vehicles and renewable energy sources on the electrical grid

[1]*Maksym OLIINYK (3rd year)*
Supervisor: [2]Jaroslav DZMURA

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

[1]maksym.oliinyk@tuke.sk, [2]jaroslav.dzmura@tuke.sk

***Abstract*—With an increase in the number of new elements in classical power supply systems, the complexity of designing or predicting the operation of networks increases. The use of individual components of the smart grid can not only reduce, but also disrupt the normal functioning of the distribution network. This leads to the need to collect, control and analyze more data that was previously impossible. The presence of a large number of measuring devices allows you to collect gigabytes of data, and every year the number of such devices will increase, which will allow energy companies to build more accurate mathematical models and make better design decisions. All of this will lead to the creation of truly smart grids.**

***Keywords*— Data analysis, electric vehicle, renewable energy, smart grids.**

I. Introduction

The power system, which is based on the Smart Grid concept ("Smart Grid", "Aquive-Adaptive Grids"), is a single complex, inside which all the elements of the electrical network are interconnected, have the ability to remotely control, control systems allow the networks to work with maximum possible efficiency. The creation of such networks allows energy supply companies to reach a completely new level of power grid management. The transition from classic electrical networks to new ones is only a matter of time. One of the main drivers of the transition will be the widespread introduction of renewable energy sources and electric vehicles. Experts point to a trend towards a rapid narrowing of the gap in the cost of conventional and non-traditional energy sources [1]. As part of the transition from classic networks to new generation networks, the European Union has identified the following important areas of development [2]:

1. Intelligent network management.
2. Demand Management (DSM).
3. Integrated production and storage (DG&S).
4. Electric transport.
5. Wide integration of renewable energy sources.

While the theoretical network modernization is simple, the smart grid deployment is a complex technical process. In this case, such instability of renewable energy sources. Energy production makes its own negative adjustments to the stable operation of power systems; modern control systems, as well as modern forecasting models (artificial intelligence), will help in the fight against these negative factors. In recent years, there has been a clear trend in the automotive industry towards a shift from classic combustion engine vehicles to electric vehicles [3]. In this regard, over the past year, in my research, I have focused on the impact of electric vehicles on power supply systems and ways to reduce their negative impact, as well as ways to predict the load in networks and generate energy from renewable energy sources. As part of these studies, several scientific articles were written. From the findings of the research, the following conclusions can be drawn [4][5]:

1. Modern networks are technically capable of withstanding the use of electrical machines up to 35% of the total number of consumers [4].

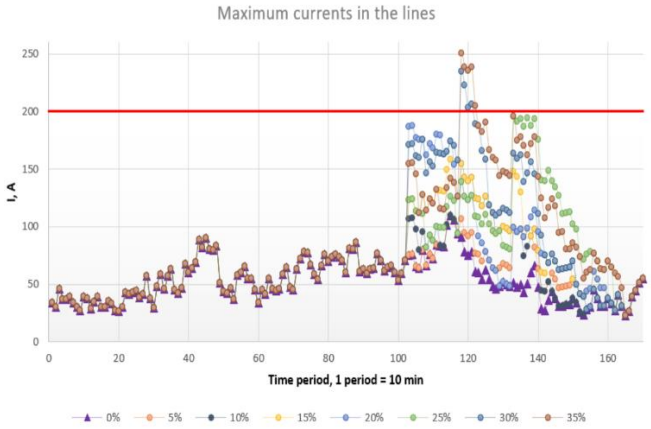


Fig. 1. Graph of maximum currents in the network[4].

2. When using electric machines by 15% of the total, losses in the electrical network increase by 100%, and when using 35%, losses increase by 330% [4].
3. The introduction of demand management systems allows to increase the number of cars up to 50-60% of the total number of consumers by distributing demand and introducing improved forecasting methods.
4. The proliferation of electric cars worsens the symmetry of the networks (single-phase charging, some electric cars use single-phase or two-phase solutions [4].
5. The widespread use of photovoltaic panels as a result of subsidies can create serious asymmetries in the network, as well as cause overvoltage's in some network nodes in good weather.

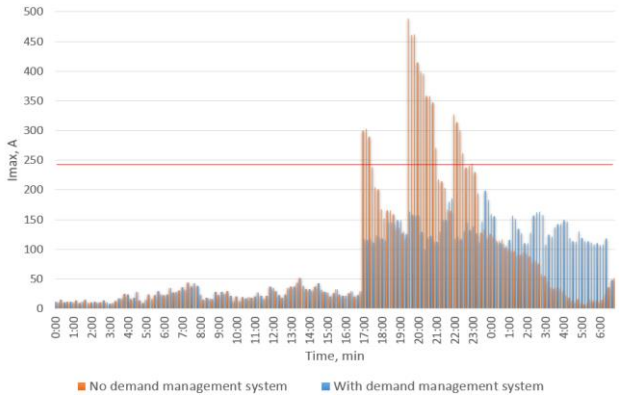


Fig. 2. Comparison of the maximum current in the network with the number of electric vehicles equal to 55%

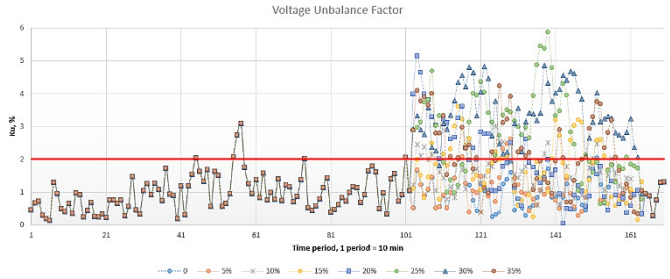


Fig. 3. Graph of voltage unbalance factor when changing the number of charging stations [4]

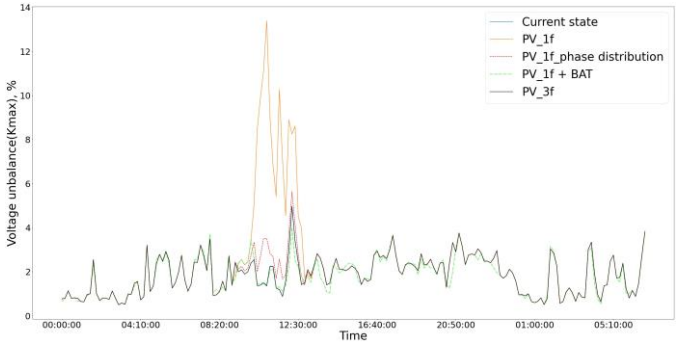


Fig. 4. Graph of voltage unbalance factor after installing 30% PV in Grid

6. Good predictive models enable better control systems in smart grids [5].

TABLE I
PREDICTION ACCURACY ASSESSMENT

Prediction	MAE	MSE	R^2
Electricity price	6.419 €/MWh	9.266 €/MWh	66,9 %
Solar power generation	11.041 MW	21.947 MW	94,1 %
Load	62.188 MW	83.815 MW	96,4 %
Cross-border power flows DE>CZ	119.534 MW	174.012 MW	82,4 %
Cross-border power flows DE>PL	124.798 MW	161.167 MW	83,6 %
Cross-border power flows DE>AT	0.41 MW	4.527 MW	99,9 %

7. A partial solution to the problems of excessive load can be solved by installing electrical energy storage systems that would consume excess during the day and reduce evening peaks.

II. Publications

To this date I've wrote and published 6 articles as 1st author and other 6 as co-author:

- AED (5),
- AFD (3)
- AFC (10),
- ADF (3).

III. Thesis of Dissertation Work

After studying the theoretical foundations about smart grids in the electric power system, a thesis exam was passed and the following PhD theses were defined:

1. Analysis of the current state of smart grids in the world
2. Technical requirements of smart grids, analysis of functionalities and their implementation in current conditions.
3. Research of mathematical models for modeling of intelligent networks.
4. Modeling of an intelligent parking house, application of RES, analysis of e-mobility usability.
5. Analysis of the achieved results and recommendations for further research.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-19-0576.

References

[1] N. Nikmehr, S. N. Ravadanegh, "Optimal power dispatch of multimicrogrids at future smart distribution grids," IEEE Transactions on Smart Grid, 2015, pp. 1648–1657.

[2] A. Mengolini, "Smart grid projects outlook 2017", Luxembourg, 2017, ISSN 1018-5593 Available on internet: <https://ses.jrc.ec.europa.eu/sites/ses.jrc.ec.europa.eu/files/u24/2017/sgp_outlook_2017-online.pdf>

[3] Global EV Outlook 2019, Available on internet: https://webstore.iea.org/global-ev-outlook-2019

[4] M. Oliinyk, J. Džmura, J. Humeník, P. Havran, R. Štefko, Data analysis and predictive models in Smart Grid- 2020. In: Elektroenergetika : International Scientific and Professional Journal on Electrical Engineering : Medzinárodný vedecký a odborný časopis pre elektroenergetiku. Roč. 13, č. 1 (2020), s. 35-39 [print]. - ISSN 1337-6756

[5] M. Oliinyk, J. Džmura, D. Pál, The impact of a electric vehicle charging on the distribution system / - 2020. In: EPE 2020: 21st International scientific conference on electric power engineering. Prague, Czech Republic. October 19-21, 2020. - Praha (Česko): České vysoké učení technické v Praze s. 18-22 [print, online]. - ISBN 978-1-7281-9479-0 - ISSN 2376-5631

Characterization and SANS study of transformer oil-based magnetic fluids

[1]*Maksym KARPETS (2nd year)*
Supervisor: [2]Milan TIMKO

[1]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1,2]Institute of Experimental Physics SAS, Košice, Slovak Republic

[1]maksym.karpets@tuke.sk, [2]timko@saske.sk

***Abstract*—This paper is a summarisational article in the range of 2 pages, which is a brief overview of my work for the past year of study. The main research task is aimed on characterization of magnetic fluids (MFs) and experimental investigation of bulk structural changes in transformer oil-based magnetic fluids (TOMFs) under the influence of electric field.**

***Keywords*—external fields, magnetic fluids, magnetic nanoparticles, transformer oil-based ferrofluids.**

I. INTRODUCTION

Magnetic fluids are suspensions of magnetic nanoparticles (MNPs) in a liquid carrier. The study of MF systems is of both fundamental and applied interest. They are widely used in various fields, in particular, they have great potential for use in engineering and medicine [1]. One of the key equipment in the electrification system is an electric transformer. During the transformer operation the heat energy is generated in the transformer windings and core. It is necessary to transfer the heat from the transformer. Recently, nanodielectric fluids have been proposed to be used as transformer liquid insulation. The addition of particles seeks to improve the dielectric and thermal properties of the base liquids, giving rise to a new generation of insulating fluids. Despite the large number of works devoted to the study of the structure of magnetic fluids in the volume of the sample [2-4], still there are properties and phenomena, the understanding of which and subsequent practical applications in electrical engineering require deeper experimental study. A mechanism of the streamer development leading to the electrical breakdown, polarization, macroscopic structural changes of magnetic fluids in external fields and nanoscopic structure at magnetic fluid-solid interface belong to the currently discussed issues. Such fields play an important role in numerous self-assembly phenomena.

II. INITIAL STATUS

With regards to the potential application of TOMFs in power transformers, most research studies have focused on three key properties: thermal, dielectric, and viscous. On the other hand, these properties are function of the base liquid, dispersed particles, size distribution, volume fraction, surfactant, and the external magnetic and electric field. The interaction of MNPs with external magnetic fields can lead to different structural re-organizations of ferrofluids including the formation of aggregates, chains and more complex patterns. Effect of an external magnetic field and the induced heterogeneous nanoparticle structure results in the magneto-viscous effect and magneto-dielectric anisotropy [2]. Also it was found that the thermal conductivities parallel to the magnetic field direction are, in most cases, much higher than the perpendicular ones [3]. At a high magnetic flux density the neutron reflectivity analysis shows the formation of two effective adsorption layers with different content of MNPs in them [4]. Thus, the magnetic field intensity, direction and distribution plays a key role in the study of MF. Analogously, structural transitions in various colloidal suspensions can be induced by electric fields leading to particle-particle electrostatic interactions. Study [5] showed that the dielectric response of MF is dependent on the strength and frequency of the electric field. It was demonstrated visually observable pattern formation in a TOMF exposed to a DC electric field. Also the electric field induced changes in the ferrofluid structure at nanoscale were recently confirmed by in situ small-angle neutron scattering (SANS) experiments [6].

III. WHAT I HAVE SOLVED THIS YEAR

A. Characterization of magnetic fluids

Preparation and characterization of magnetic fluids is the first step in their further investigations. The several new types of transformer oil (TO) were used as a carrier liquid in the studied ferrofluid. Iron oxide MNPs were synthesized by chemical co-precipitation method from aqueous solution of ferrous and ferric ions in the presence of NH_4OH at 80–82 °C. After coprecipitation, the NPs were sterically stabilized by chemisorbing of a single oleic acid layer ($C_{18}H_{34}O_2$, Merk) on the particle surfaces. Synthesis of MF was performed according to previous reports [4-6]. The concentration of the NPs in the MF was chosen to be from 0.6 to 2.8 vol %.

The magnetic properties of the powder MNPs and MF samples were measured by means of a vibrating sample magnetometer installed on a cryogen-free superconducting magnet from Cryogenic Limited (IEP SAS, Kosice). The mean particle size derived from the fitting of the magnetization curve by the superposition of Langevin functions is 9 nm. From inset 2 of Fig. 1, one can observe the typical superparamagnetic behavior with zero coercivity and remanence. At higher magnetic fields (above 2 T), the magnetization of MF is well saturated. The mean hydrodynamic particle diameter

determined from dynamic light scattering is about 19 nm.

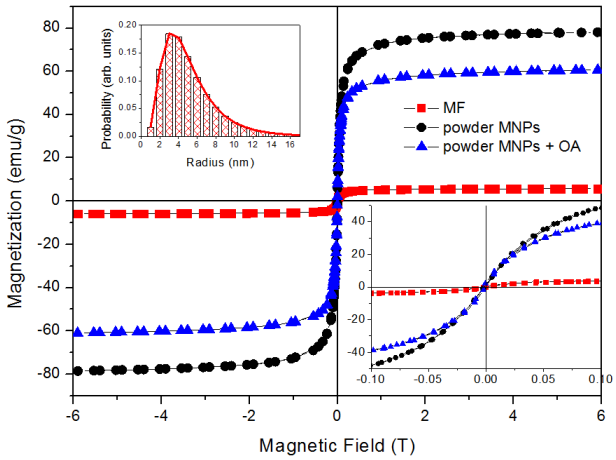


Fig. 1. Magnetization curves of the powder magnetite NPs, NPs with surfactant and MF. Inset 1 (left upper corner): distribution of MNPs in MF calculated from the magnetization. Inset 2 (right bottom corner): superparamagnetic behavior (zero hysteresis) of the samples.

To confirm the considered lossy dielectric response of the MF samples, we present the dielectric loss of MF samples and the TO in the form of temperature-dependent dissipation factor measured at 1 kV of the measuring AC voltage with the frequency of 50 Hz using Eltel ADTR-2K [7].

Not only magnetite but Mn–Zn ferrite nanoparticles based magnetic fluids were characterized. The AC magnetic susceptibility was measured at room temperature by AC susceptometer Imego Dynomag, in the excitation field with the amplitude of 0.5 mT. The nearly constant behavior reflects the absence of any particle aggregates that would manifest themselves by a rapid decrease of the real susceptibility. Dielectric and insulating properties of the oils and the nanofluids was investigated at IEP SAS in the Laboratory of dielectric measurements by dielectric spectroscopy method – LCR meters with the frequency range from 0.1 Hz to 200 kHz and from 20 Hz to 2 MHz (Agilent E4980A). More concentrated samples yield higher value of the frequency-dependent real and imaginary parts of complex dielectric permittivity.

B. SANS study of magnetic fluids under electric field

In situ SANS experiments were carried out on the small-angle neutron diffractometer - time-of-flight YuMO instrument at the pulsed IBR-2 reactor at JINR, Russia [8]. All received SANS data could be well treated with theoretical model of core-shell particles. The obtained experimental SANS dependencies and fit for the ferrofluids based on insulating transformer oils SHELL and MOL with different nanoparticles concentration for voltage 0 kV are plotted in Fig. 2. One can see significant increasing in scattering intensity with increasing of iron oxide NPs concentration. The cause of it is more quantity of scattering centers. Similar behavior is confirmed also in presence of electric field. But actually there are almost no changes in SANS data with applying and increasing DC electric field strengths for all samples. Therefore expected aggregates are probably of greater size which is not possible to see in the q-range covered in the experiment. The mean radius of NPs calculated from fit was about 4 nm.

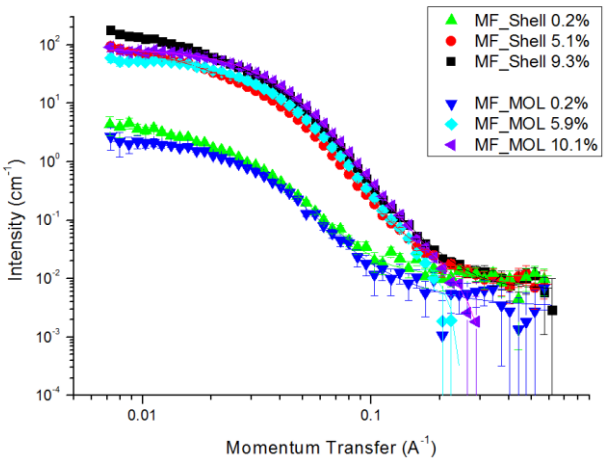


Fig. 2. SANS data (points) and fit (curves) for the ferrofluids based on insulating transformer oils SHELL and MOL with different magnetic mass fraction of NPs.

IV. Conclusion and Next Steps

Magnetic and dielectric properties of prepared MFs were characterized by several methods. We have demonstrated NPs concentration influence on SANS measurements of the classical MFs. As the next step we will try to investigate the temperature effect on structure of TOMF under electric field by SAXS as well as the magnetic fluid-solid interfaces under the influence of magnetic field by neutron reflectometry. The experimental proposals were already accepted for measurements and will be granted with beam time in 2021 at BioSAXS (Hamburg, Germany) and JINR (Dubna, Russia) accordingly. We are also interested in the implementation of fullerene NPs in our MF for structural investigation of hybrid magnetic nanofluids at external fields. Extremely important will be study of combine effect of both electric and magnetic field on structure of MFs.

Acknowledgment

The work is a part of the projects supported by the Programme of SAS grants for PhD students (No. APP0140); the Science Grant Agency of the Slovak Republic (No. 2/0011/20 and 2/0016/17); the Slovak Research and Development Agency (No. APVV-18-0160).

References

[1] S. Odenbach, Colloidal magnetic fluids: basics, development and application of ferrofluids, Vol. 763, Springer Berlin Heidelberg, (2009).
[2] M. Rajnak et al., "Magnetic field effect on thermal, dielectric, and viscous properties of a transformer oil-based magnetic nanofluid", Energies 12, 4532 (2019).
[3] A. Katiyar, P. Dhar, T. Nandi, and S. K. Das, "Magnetic field induced augmented thermal conduction phenomenon in magneto-nanocolloids", J. Magn. Magn. Mater. 419, 588 (2016).
[4] A. Nagornyi et al., "Particle assembling induced by non-homogeneous magnetic field at transformer oil-based ferrofluid/silicon crystal interface by neutron reflectometry", Appl. Surf. Sci. 473, 912 (2019).
[5] M. Rajnak et al., "Dielectric response of transformer oil based ferrofluid in low frequency range", J. Appl. Phys., 114, 3 (2013).
[6] M. Rajnak et al., "Structure and viscosity of a transformer oil-based ferrofluid under an external electric field", J. Magn. Magn. Mater. 431 99–102 (2017).
[7] M. Rajňák et al., "Statistical analysis of AC dielectric breakdown in transformer oil-based magnetic nanofluids", J. Mol. Liq. 309, 113243 (2020).
[8] M. Karpets et al., "Small-angle neutron scattering study of transformer oil-based ferrofluids", Ukr. J. Phys. 65, 729 (2020).

Methods and algorithms for acceleration / parallelization of physical model calculations

[1]*Michal SOLANIK (1st 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 SAV Kosice, Slovak Republic

[1]michal.solanik@tuke.sk, [2]jan.genci@tuke.sk, [3]bobik@saske.sk

***Abstract*—Implementation of Parker's transport equation stochastic solution for simulations of cosmic rays distribution in heliosphere can be demanding on computer performance. GPU accelerated implementation however requires a decrease of accuracy against CPU implementation. In this paper, we will discuss the available parallel computing platform used by Monte Carlo simulations and aim for the current state of acceleration of cosmic rays distribution simulations in the heliosphere.**

***Keywords*—Heliosphere, GPGPU, CUDA, Distributed systems**

I. Introduction

Parker's transport equation can be solved [1] with the stochastic approach in several ways:

- B-p - Backward-in-time with momentum p
- B-T - Backward-in-time with kinetic energy T
- F-p - Forward-in-time with momentum p
- F-T - Forward-in-time with kinetic energy T

These solutions are demanding computing resources and require days or even weeks to complete simulations on standard CPUs. In this paper, we will discuss available platforms, GPU implementation backward-in-time and forward-in-time methods with possible impacts on the overall accuracy of the model, and existing Monte Carlo simulation type optimizations.

Many authors [2] [3] [4] [5] tried to gain acceleration by creating GPU implementation of Monte Carlo models. Acquired acceleration usually ranged from 3 to 10 while depending on the type of model. However, in specific cases, even speedup up to 60 was [2] possible due to the nature of the given model.

Keeping accuracy at a sufficient level is key to use these models with real data. HelMod model [6] verified its functionality against measured data from probes. Implementing backward or forward model on GPU [2] require using single-precision floating-point format (we will refer to this as FP32) instead of double-precision float-point format (we will refer to this as FP64). We will discuss this topic in details in section III.

II. Computing platforms used in Monte Carlo simulations

Monte Carlo is [7] stochastic method that use random numbers to obtain results from simulations. Forward-in-time and backward-in-time heliospheric simulations are time-consuming. These simulations with a particular range of input parameters can last weeks or even months on single standard CPUs. We will take a look at other platforms used in scientific Monte Carlo simulations besides standard CPUs.

A. GPGPU

GPGPU (General-purpose computing on graphics processing units) [8] is utilization of GPU as a computation unit for general-purpose computing. GPU consists of multiple multiprocessors that are less complex than the standard CPU. One of the main differences between CPU and GPU is thread scheduling. On CPU host operating system [8] manages threads and their execution, but on GPU threads are managed by the scheduler.

Three parallel frameworks are usually used [9] for GPU programming:

- CUDA - framework that allow use [8] CUDA based GPUs for GPGPU, exclusively for Nvidia GPUs.
- OpenCL - a framework determined for writing programs that can run on multiple types of devices.
- OpenACC - [10] rely on compiler directives API in C/C++ and Fortran. It is able to parallelize sequence code using compiler directives.

Authors Memeti et al. [9] discussed efficiency of these frameworks from a productive and performance perspective. OpenACC implementations were the slowest from performance aspects, but the most productive in the creation of code compared to OpenCL and CUDA. This is not unexpected as OpenACC [10] aims on fast and simple parallelization of sequential code. Comparison between OpenCL and CUDA resulted in more than half items differences were at the level of 20% of execution time but in the case of Kmeans, Heartwall, and LavaMD algorithms results showed more than 50% differences in execution time. The authors did not explain [9] possible reasons for these differences. Authors Karimi et al. compared [11] OpenCL and CUDA implementations on GPU for Monte Carlo simulation of Adiabatic QUantum Algorithms. CUDA implementation was 14.54% faster.

GPGPU is also used even in the heliospheric field. Authors Dunzlaff et al. [2] developed GPU implementation of Parker's transport equation solution for energetic charged particles with 10-60x speed-up against reference system. In section III we will discuss their work in more detail. In [12] we parallelized implementation of 1D Parker's transport equation solution on

GPU with gained speed-up ranged from 7.71-86.87x against reference CPU system.

Even outside heliospheric field multiple authors [13] [14] [15] achieved results that GPU implementation (written in OpenCL or CUDA) of any Monte Carlo simulation was significantly faster than CPU implementation.

B. FPGA

FPGA (Field-programmable gate array) [16] is an integrated circuit that contains high-density arrays that can execute a program in parallel. Due to its parallel nature, it can be used as an alternative to the GPU. In heliospheric field FPGAs [17] [18] are usually used in measuring instruments. Several authors implemented [19] [20] [21] solution for their Monte Carlo simulations with even successful gain against single GPU.

Implementation of Monte Carlo simulations for the Ising model on FPGA authors in [19] reported significant 11.8x acceleration against single GPU and in [21] even 2.98x speed-up against GPU cluster that was based on [19]. Authors Lin et al. [19] did not mention type of reference GPU or type of GPU used in GPU cluster. Gained acceleration against the GPU cluster is the result of technological advances in the FPGA field. But also the result of comparison between hardware with a significant time gap.

Tian et al. compared [20] CPU, GPU, and FPGA implementation of quasi-monte Carlo financial simulation. According to the authors, FPGA is an alternative to GPU as they gained a 3x acceleration with an FPGA-based cluster against GPU. However, their comparison of the performance of the entire cluster against a single GPU is not correct. Price per GFLOP is usually higher in FPGAs than GPUs.

Using FPGA as computing platform can be an alternative against using GPU. We can gain significant acceleration against conventional CPUs, but the high price per GFLOP against GPUs is negative.

C. Distributed systems

Distributed systems [22] are usually defined as a group of independent computers that act as one system from the user's view.

We can divide distributed systems by composition of computing units to:

- Homogeneous distributed systems - each computing unit is of the same type,
- Heterogeneous distributed systems - computing units in a distributed system can be of a different type.

Several authors [23] [24] tried to solve their Monte Carlo simulation problem on distributed systems. Implementation of Parker's transport equation stochastic solution on the heterogeneous distributed system using GPU as a computing unit [25] proved to be useful. We will go in more details in section V.

Camarsu-Pop et al. [23] implemented framework based on merging partial results from each computing unit. This led to a reduction of execution time by 40%. The authors claim that they used a heterogeneous distributed system consisting of CPUs.

Szałkowski et al. [24] used GPU-hybrid cluster for testing implementation of the linear congruential generator and lagged Fibonacci generator in Monte Carlo simulation.

Both authors [23] [24] came to the conclusion that distributed systems are suitable for Monte Carlo simulations and can offer additional speed-up against the non-distributed system. Using multiple GPUs or FPGAs in a distributed system can also enhance performance against a classical distributed system at the cost of more complex control over the simulation.

III. Parallelization of heliospheric models

Dunzlaff et al. [2] compared GPU and CPU versions of solving Parker's transport equation by stochastic differential equations. Increased performance ranged from 10 to 60 depending on input parameters. They focused on random number generation, fast-math comparison, thread divergence and gained speedup against a reference computer system.

GPU implementation [2] used XORWOW random number generator. However, documentation for CURAND suggests [26] that this type of random number generator has issues with distribution. Saito and Matsumoto analyzed [27] failed tests in BigCrush test framework. They suggested that only one of three failed tests is crucial for the Monte Carlo type of simulation. The failure occurred on collision over test on the 7-dimensional distribution with deviation from desired distribution. Multiple authors referred [4] [3] [5] to this issue but none of them measured or mentioned impact of possible deviation in their simulations. Dunzlaff and et al. compared [2] GPU implementation using XORWOW and MTGP32 random number generators. After injecting approximately 400 000 particles for both GPU implementations and another CPU as reference version they received very similar results without any observable deviations. The absence of deviations can be explained by two hypotheses:

- used models are not affected by minor deviation in used distribution,
- execution time led to the situation that distribution deviation could not affect the simulation.

The second hypothesis can be confirmed, as a test where deviation in distribution [26] was found lasted nearly six hours of generating random numbers. In [25] we also compared output spectra from GPU implementation of 1D forward-in-time model with XORWOW and MTGP32 random number generators but without any visible deviations or inaccuracy.

GPU thread scheduling is managed by GPU and not by an operating system as in the case of CPU which we mentioned in section II-A. This can lead to thread divergence. To minimize thread divergence authors in [2] divided simulation into partitions with the predefined number of steps. The authors observed that the most efficient number of partitions is 60-170 with the best result achieved with 90 partitions. However, this approach can be difficult for multiple edge cases of input parameters into the simulation. As for this, they would need at least a function of input parameters that can calculate the required numbers of steps for a given combination.

Additional optimization can be done on the compiler level - usually sacrificing accuracy in favor of speed. In this case, the authors [2] compared version with the usage of fast math optimization against usage of O3 optimization. The authors did not mention used simulation type. However, we assume that backward-in-time was used because of the used injection point for pseudo-particles. On 1 AU (Astronomical unit) observed maximum deviation was 5%. We [12] tested 1D backward-in-time and also forward-in-time with fast math on GPU. In

the case of 1D backward-in-time, no significant deviations were observed, however in the case of 1D forward-in-time we observed two visible differences with time step 0.5s.

IV. Accuracy of Heliospheric Models

Bobik et al. [1] compared differences between forward-in-time models and backward-in-time models against the Crank-Nicolson model. Backward-in-time models showed a maximum deviation of 1% at low energies. The maximum deviation in the case of F-p and F-T models ranged from 5% to 10%. The lowest energy displayed in the chart is 0.5 GeV so it is likely that the deviation between F-T and CN is more than 10% on energies lower than 0.5 GeV.

We started the discussion about the accuracy of heliospheric models in section III. Dunzlaff et al. [2] had no issues with the accuracy of their GPU implementation for forward-in-time or backward-in-time model. In 1D forward-in-time model we [25] found defect that was no present in backward-in-time model with same input parameter, time step dt < 2.0 s and [1] was not also described by Bobik et al..

Figure 1 shows ratio between 1D F-p and Crank-Nicolson model energetic spectra. Spectrum in the GPU model has several unexpected peaks that are evenly found on the entire spectrum range. We suggested [25] several hypotheses that can cause these pulsations:

- used random number generator,
- used *–use_fast_math* parameter,
- incorrect distribution of injected energies,
- differences between single and double precision.

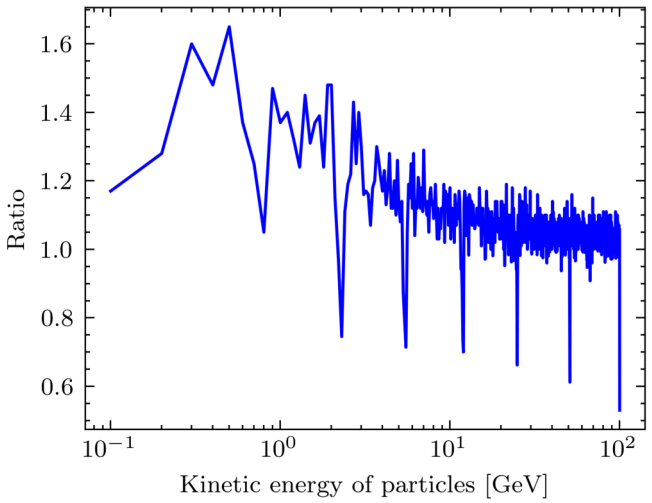


Fig. 1. Energetic spectrum for 200 bilions of particles - F-p model - dt = 0.5

As we already mentioned in section III XORWOW random number generator [27] failed in a few tests of BigCrush test framework. Usage of MTGP32 random number generator [25] did not help and pulses remained in same locations.

In [12] [25] we used *–use_fast_math* parameter as authors Dunzlaff et al. did [2]. Usage of this parameter had no significant impact on overall accuracy, as mentioned in section III. In figure 2 is shown energetic spectrum without usage of *–use_fast_math* parameter. Is clear, that the *–use_fast_math* parameter is not the reason for pulsations, but caused deviations on lower energies. In the version with *–use_fast_math* parameter first visible pulse is between 2 and 3 GeV. The first pulse is visible with its peak on 0.6 GeV without the usage of fast math.

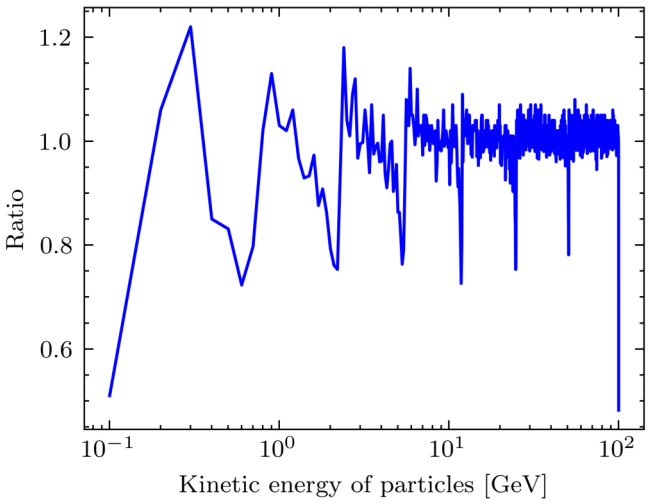


Fig. 2. : Energetic spectrum for 200 bilions of particles - F-p model - dt = 1.5 without usage of *–use_fast_math*

Incorrect distribution of injected energies hypothesis was not confirmed [25] as output spectrum remained same as is shown in figure 1. The last hypothesis with differences between single and double precision has not been tested yet.

V. Heterogeneous Distributed System for Cosmic Rays Distribution Models

Abbreviated execution time on a single GPU was [25] impulse to design heterogeneous distributed system using multiple GPUs.

Figure 3 shows architecture of heterogeneous distributed system. The system contains user interface for launching new simulations, administration of all simulations, and tools for rendering energetic spectrum. The primary node of distributed system distributes simulations for each available node of a distributed system via a message broker.

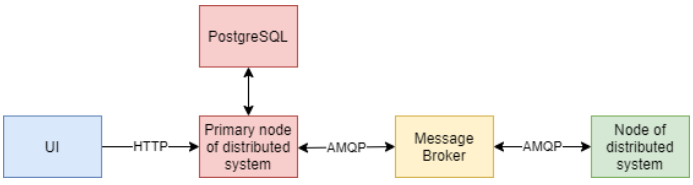


Fig. 3. Archicture of heterogeneous distributed system for cosmic rays distribution models

The most significant issues of the distributed system were how to run simulations on the node, how to distribute simulations per node in the most equal way, and how to design and implement fault-tolerance.

The node of the distributed system is waiting for the message with details about the new simulation. After receiving a new message simulation will be launched with given parameters. In [25] we decided to use a GPU application with a command-line interface to launch simulation on the node of the distributed system for more simple maintenance.

Usage of multiple different GPUs complicated the distribution of simulations across systems. For backward-type models, equal distribution per node in a test environment consisting

TABLE I
EXECUTION TIMES FOR EACH MODEL TESTED IN DISTRIBUTED SYSTEM

System type	1D F-p [h]	2D F-p [h]	1D B-p [h]
Reference system	43.30	150.00	145.57
GTX 1080 TI	5.61	12.82	1.66
Distributed system	4.28	10.50	1.17

of RTX2060 and GTX1080TI proved to be efficient. This approach was ineffective in the case of the forward-type models. We [25] used different approach using simple benchmark. Benchmark value was based on GPU FP32 performance, memory bandwidth, and usage of GPU for a given model. Execution time in tested distributed system decreased from 19.89% to 34.62% against usage of single GTX1080TI as is shown in table I.

The risk of failure in the distributed system [22] is multiplied by using more and more nodes. In our distributed system in [25] we implemented simple regular message that is sent from node to the primary node. If one of the nodes is not available for 15 minutes its current task is distributed to another node. However, this is done regardless of the efficiency that can be achieved in the current status of the distributed system.

VI. FURTHER RESEARCH DIRECTIONS

Using GPU as a computing platform for simulation Parker's transport equation stochastic solution [12] [25] proved to be suitable with acceptable results inaccuracy. Extension to a distributed system using GPU as a computing unit is also a suitable platform for solving Monte Carlo simulations.

In our future research, we want to focus on distributed systems and their effectiveness across multiple computing platforms. Supporting combined multiple CPU and GPU computation or even FPGA can also reduce overall execution time for heliospheric models.

Parallelization of Monte Carlo simulations on FPGA [19] [21] proved to be an alternative to graphic processing units as well. Compared to GPUs, FPGAs have one substantial drawback, cost per GFLOP. However, FPGAs can serve as a complement to GPU.

Accuracy of GPU implementation is affected [25] by pulses with time step dt < 2.0 s. After a more detailed analysis, we should be able to identify problematic parameters. These simulations can be calculated on CPU with FP64 precision and their results fused with results from GPU. Elimination of pulsation along with the preservation of gained acceleration can be achieved in a distributed system with a sufficient amount of nodes with CPU and GPU nodes.

Heterogeneous distributed system performance has the potential to increase, especially in scheduling. The new scheduler should be more oriented on shorting execution time and starting new simulations concerning already occupied nodes.

REFERENCES

[1] P. Bobik, M. Boschini, S. Della Torre, M. Gervasi, D. Grandi, G. La Vacca, S. Pensotti, M. Putis, P. Rancoita, D. Rozza *et al.*, "On the forward-backward-in-time approach for monte carlo solution of parker's transport equation: One-dimensional case," *Journal of Geophysical Research: Space Physics*, vol. 121, no. 5, pp. 3920–3930, 2016.

[2] P. Dunzlaff, R. Strauss, and M. Potgieter, "Solving parker's transport equation with stochastic differential equations on gpus," *Computer Physics Communications*, vol. 192, pp. 156–165, 2015.

[3] X. Jia, H. Yan, X. Gu, and S. B. Jiang, "Fast monte carlo simulation for patient-specific ct/cbct imaging dose calculation," *Physics in Medicine & Biology*, vol. 57, no. 3, p. 577, 2012.

[4] X. Jia, J. Schümann, H. Paganetti, and S. B. Jiang, "Gpu-based fast monte carlo dose calculation for proton therapy," *Physics in Medicine & Biology*, vol. 57, no. 23, p. 7783, 2012.

[5] L. Xu and G. Ökten, "High-performance financial simulation using randomized quasi-monte carlo methods," *Quantitative Finance*, vol. 15, no. 8, pp. 1425–1436, 2015.

[6] M. Boschini, S. Della Torre, M. Gervasi, G. La Vacca, and P. Rancoita, "Propagation of cosmic rays in heliosphere: The helmod model," *Advances in Space Research*, vol. 62, no. 10, pp. 2859–2879, 2018.

[7] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American statistical association*, vol. 44, no. 247, pp. 335–341, 1949.

[8] "Cuda toolkit documentation v11.2.0." [Online]. Available: https://docs.nvidia.com/cuda/

[9] S. Memeti, L. Li, S. Pllana, J. Kołodziej, and C. Kessler, "Benchmarking opencl, openacc, openmp, and cuda: programming productivity, performance, and energy consumption," in *Proceedings of the 2017 Workshop on Adaptive Resource Management and Scheduling for Cloud Computing*, 2017, pp. 1–6.

[10] S. Wienke, P. Springer, C. Terboven, and D. an Mey, "Openacc—first experiences with real-world applications," in *European Conference on Parallel Processing*. Springer, 2012, pp. 859–870.

[11] K. Karimi, N. G. Dickson, and F. Hamze, "A performance comparison of cuda and opencl," *arXiv preprint arXiv:1005.2581*, 2010.

[12] M. Solanik, "Model of cosmic rays distribution in heliosphere redisgn to cuda language," 2018.

[13] J. Spiechowicz, M. Kostur, and L. Machura, "Gpu accelerated monte carlo simulation of brownian motors dynamics with cuda," *Computer Physics Communications*, vol. 191, pp. 140–149, 2015.

[14] M. Goldsworthy, "A gpu–cuda based direct simulation monte carlo algorithm for real gas flows," *Computers & Fluids*, vol. 94, pp. 58–68, 2014.

[15] N. Henderson, K. Murakami, K. Amako, M. Asai, T. Aso, A. Dotti, A. Kimura, M. Gerritsen, H. Kurashige, J. Perl *et al.*, "A cuda monte carlo simulator for radiation therapy dosimetry based on geant4," in *SNA+ MC 2013-Joint International Conference on Supercomputing in Nuclear Applications+ Monte Carlo*. EDP Sciences, 2014, p. 04204.

[16] T. A. York, "Survey of field programmable logic devices," *Microprocessors and Microsystems*, vol. 17, no. 7, pp. 371 – 381, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/014193319390059G

[17] R. A. Howard, A. Vourlidas, C. M. Korendyke, S. P. Plunkett, M. T. Carter, D. Wang, N. Rich, D. R. McMullin, S. Lynch, A. Thurn *et al.*, "The solar and heliospheric imager (solohi) instrument for the solar orbiter mission," in *Solar Physics and Space Weather Instrumentation V*, vol. 8862. International Society for Optics and Photonics, 2013, p. 88620H.

[18] N. Ampilogov, I. Astapov, N. Barbashina, V. Borog, D. Chernov, A. Dmitrieva, K. Kompaniets, A. Petrukhin, V. Shutenko, A. Teregulov *et al.*, "Large area scintillation muon hodoscope for monitoring of atmospheric and heliospheric processes," *Astrophysics and Space Sciences Transactions*, vol. 7, no. 3, pp. 435–438, 2011.

[19] Y. Lin, F. Wang, X. Zheng, H. Gao, and L. Zhang, "Monte carlo simulation of the ising model on fpga," *Journal of computational Physics*, vol. 237, pp. 224–234, 2013.

[20] X. Tian and K. Benkrid, "High-performance quasi-monte carlo financial simulation: Fpga vs. gpp vs. gpu," *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, vol. 3, no. 4, pp. 1–22, 2010.

[21] F. Ortega-Zamorano, M. A. Montemurro, S. A. Cannas, J. M. Jerez, and L. Franco, "Fpga hardware acceleration of monte carlo simulations for the ising model," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 9, pp. 2618–2627, 2015.

[22] A. S. Tanenbaum and M. Van Steen, *Distributed systems: principles and paradigms*. Prentice-hall, 2007.

[23] S. Camarasu-Pop, T. Glatard, R. F. Da Silva, P. Gueth, D. Sarrut, and H. Benoit-Cattin, "Monte carlo simulation on heterogeneous distributed systems: A computing framework with parallel merging and checkpointing strategies," *Future Generation Computer Systems*, vol. 29, no. 3, pp. 728–738, 2013.

[24] D. Szałkowski and P. Stpiczyński, "Using distributed memory parallel computers and gpu clusters for multidimensional monte carlo integration," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 4, pp. 923–936, 2015.

[25] M. Solanik, "Paralelization of two-dimensional models of cosmic rays distribution in heliosphere," 2020.

[26] "Testing." [Online]. Available: https://docs.nvidia.com/cuda/curand/testing.html#testing

[27] M. Saito and M. Matsumoto, "A deviation of curand: standard pseudo-random number generator in cuda for gpgpu," 2012.

Controlling SMART devices via unified user interface in Virtual Reality environment

[1]*Martin SIVÝ (4th year)*
Supervisor: [2]Branislav SOBOTA

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

[1]martin.sivy@tuke.sk, [2]branislav.sobota@tuke.sk

***Abstract*—This paper describes the use of a proposed unified user interface in virtual reality for the purpose of controlling any SMART device located in a SMART household. It focuses on the creation of a web-based cross-platform client-side application that is able to manage a user's household appliances in a unified way that takes into consideration the specific attributes or characteristics set by the manufacturer of such device. This application is able to connect any device to any SMART household, generate its virtualized controls according to predefined attributes set by the manufacturer, and is able to manage requests generated by user's interaction via a centralized server. The application can manage real-time households or create fully simulated environments. This approach can be used in order to save resources while creating and designing fully simulated SMART environments with the emphasis on fast prototyping and a high level of flexibility, omitting multitudes of different hardware controllers used in households that are needed to operate every single item in the environment, or creating safe simulated environments used for training or teaching in high-risk areas of the industry. This proposed application was tested in the LIRKIS laboratory due to the availability of suitable virtualization devices such as Oculus Quest of Microsoft Hololens.**

***Keywords*—virtual reality, user interface, simulation, smart environment.**

I. Introduction

This paper describes the creation and use of web-based cross-platform client-side application for management of various household devices in virtual reality regardless of their type, manufacturer or its controllable features. It provides easy to use unified user interfaces that are represented in the form of control menus. These controllers are created according to device-specific data provided by the manufacturer in form of raw JSON data that are stored and used by the centralized server. Application takes into account any future device that could be connected to the household and is able to generate virtualized control menus just by reading the device's attributes.

The main emphasis is laid on the interaction between the user and the virtual environment. It was necessary to create a unified user interface since the variability of hardware controllers and device characteristics are endless. Our work was focused on visualizing controllers in a virtual SMART environment that unifies the way that the user controls his household. He doesn't need to use multiple different hardware controllers that would be carried around and used for the management of devices located in one household. All he has to do is connect his devices to the household and get familiar with the singular unified user interface provided by the application.

It was necessary to find a correct way to virtualize such controllers, determine what controls will be available, and specify in what way the user will interact with the environment. The user's movement in a simulated environment is represented by his avatar through which he is able to interact with real-life or simulated devices. The current state of the devices is stored and managed by the server that communicates with available devices, performs operations on real-life appliances, or fully simulates their state in the case of a simulated environment. The server also generates and governs virtualized controllers that are styled and displayed by the application.

The proposed approach can be used not only for designing and managing households but also for training and education purposes in multiple fields of the industry. Training using simulated virtual reality can be helpful in recreating life-threatening situations in a work environment utilizing the safety of virtualization, e.g. to prepare medical staff and army forces for difficult life situations. Using virtual reality in this way can yield better results in regards to memorability, engagement, or immersion compared to older or outdated approaches such as learning only from the book [1].

Different approaches for the creation of user interfaces were examined in the work of A.Lis-Marciniak et al. [2] who researched three different control mechanics in a tower defense game designed for mixed reality. In essence, they compared three different approaches for user interaction with the environment differentiating in multiple factors. Their best rated approach was also used in work proposed by this paper, such as displaying the control menu right next to the device in a fixed position regardless of where the user stands until the user decides to close it. Our solution is web-based, cross-platform, and has been designed and implemented using A-frames. It can be run on any user device that has a web browser and is able to connect to the internet and is able to provide device orientation and device motion data.

II. Interaction with the unified interface

There are several approaches on how to control objects and avatars in virtual space. This section describes techniques used in our application.

A. Gaze

Gazing (as in focusing on an item for a longer period of time) in virtual reality in order to trigger user interaction is the most common approach of control in simulated environments [3]. It is used in headsets that lack remote controls, or virtual environments visualized by mobile phones or computer screens, as well as in devices supporting 3DoF tracking. The basis for interacting with objects is a cursor whose position is tied to the position of the camera. In A-Frame, the cursor uses the principle of the raycaster method [4]. Its starting point is the camera, and the direction is its rotation. It is also possible to add a raycaster component to any entity and define its attributes, such as the targeting time required to execute a click event or types of objects it can interact with. Different behaviors are specified in our solution such as controlling device attributes using a slider, when operations are executed for the whole duration of gaze when sliding the control on a bar, or they can be executed once per gaze such as clicking a simple button.

B. Voice

Voice control as an addition to gaze, although perhaps not fully efficient on its own in virtual reality, can improve the usability of the proposed system. A-frame itself does not contain a component that can respond to the user's voice, so aframe-speech-command-component has been created. It uses the annyang library [5] to recognize a voice command in order to perform a defined action. A disadvantage of this solution is that annyang is only compatible with Google Chrome. Voice recognition is used for commands that are not device specific such as showing the global menu by saying SHOW MENU. Hints with supported voice commands are located in the upper left corner of the user screen.

C. Supported controller elements

Every virtual controller follows a default layout that consists of the device name and various actions that control how the window is displayed (such as zoom or button for close action) and a styled interactive form which consists of elements that represent controllable device attributes. Supported elements are displayed in Fig. 1 and have to be specified and provided by the manufacturer. Our approach supports buttons, labels, images, toggles, sliders, radio buttons, checkboxes, inputs, pluses, minuses and timers. Actions that take place after interacting with these controls are run according to the behavior specified by the manufacturer.

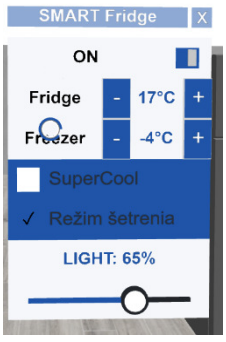


Fig. 1 Demo of appliance interface

III. CLIENT APPLICATION

Users can open our application by navigating to a web page in any browser. User can register, login in or retrieve their lost password on the landing page. After logging in, he will see his virtual household where he can interact with appliances, the example is shown in Fig. 2. Interactions are triggered by his gaze, which is indicated by changing the cursor to a loader, after which the menu of an appliance is displayed consisting of individual supported elements. All requests for changes are sent to the server that keeps the virtual household up to date with other household members that are logged in. Changes are shared using WebSocket.

Fig. 2 Prototype of kitchen with appliances and open control menu

IV. CONCLUSION

We introduced a unified user interface for virtual reality with the aim to control real life or simulated devices in smart environments regardless of their type or manufacturer. We demonstrated examples of generic virtualized appliance controllers to simplify user interaction with his environment.

Our solution can save costly resources by providing a fast and flexible method for prototyping, designing households or creating safe environments for simulating life-threatening situations valuable in training in high-risk areas of industry.

We would like to focus on testing the proposed environment with users and gain their feedback to improve the interface in the future. Demo application with bypass login can be found on this path https://spiritual-natural-yarrow.glitch.me

ACKNOWLEDGMENT

The This work has been supported by the APVV grant no. APVV-16-0202 "Enhancing cognition and motor rehabilitation using mixed reality" and by the KEGA grant No. 035TUKE-4/2019: „Virtual-reality technologies and handicapped people education".

REFERENCES

[1] K. Ahir et al. Application on Virtual Reality for Enhanced Education Learning, Military Training and Sports. In: Augmented Human Research, vol. 5, pages 1-9. 2020

[2] A. Lis-Marciniak, J. Tomiakowski and P. Kapusta. Design Rules, Implementation and Testing of User Interfaces for Mixed Reality Applications. In: 2018 11th International Conference on Human System Interaction (HSI). pages 345-351. IEEE, 2018

[3] MIT. Dokumentácia A-Frame [online]. 2020 [cit. 2021-02-07]. Dostupné z: https://aframe.io/docs/1.0.0/introduction/.

[4] PIETROSZEK, Krzysztof. Raycasting in Virtual Reality. In: Encyclopedia of Computer Graphics and Games. Springer International Publishing, 2018, s. 1– 3. ISBN 978-3-319-08234-9. Dostupné z DOI: 10.1007/978-3-319-08234- 9_180-1.

[5] ATER, Tal. Annyang [https://github.com/TalAter/annyang]. GitHub, 2021.

Robust data transmission in 5G networks without infrastructure

[1]*Maroš BAUMGARTNER (1st year)*
Supervisor: [2]Jozef JUHÁR

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]maros.baumgartner@tuke.sk, [2]jozef.juhar@tuke.sk

***Abstract*— Nowadays, the requirements for robustness, speed and efficiency of data transfer are increasing. These and many other requirements can be addressed by fifth generation networks. One of the main requirements for 5G networks is robustness, so in this article we deal with data transmission using networks without infrastructure using alternative techniques for robust data transmission.**

***Keywords*—mobile ad hoc network, 5G network, blockchain, machine learning, network simulator 3**

I. Introduction

Wireless mobile networks and data transmission using them are among the rapidly evolving field in the field of telecommunications. Fifth generation (5G) networks are addressing ever-increasing demands for higher data rates, better connection availability and data security.

5G networks offer a connection of several wireless technologies, e.g., software defined radio, mobile ad hoc networks, cognitive radio and so on.

Mobile ad hoc networks (MANET) operate without an infrastructure, where each node in the network acts as a router. With such uncontrolled communication and the connection of new unverified nodes, several security issues arise, especially during the routing process. Blockchain technology can be the answer to the question of increasing the security of decentralized networks. Blockchain is a database that integrates decentralization, distributed computing, asymmetric encryption, timestamp and consensus algorithm. The principle of blockchain lies in the continuity of individual blocks of data on each other, which prevents any intervention and modification of already created data blocks.

II. Mobile Ad Hoc network

The MANET network (Mobile Ad Hoc Network) is an autonomous system of mobile nodes. It is a wireless network that consists of mobile platforms, referred to as nodes, which can be statically in one place or can move freely. We can consider end terminals as devices that have wireless technology, such as mobile phones, PDAs, tablets or laptops. A MANET network with terminals can be described as an unoriented graph, the vertices of which are the terminals and the edges representing the wireless connection of two terminals that are within radio range. An example of a MANET network topology with an undirected graph representing a MANET network can be seen in Fig. 1 and Fig. 2.

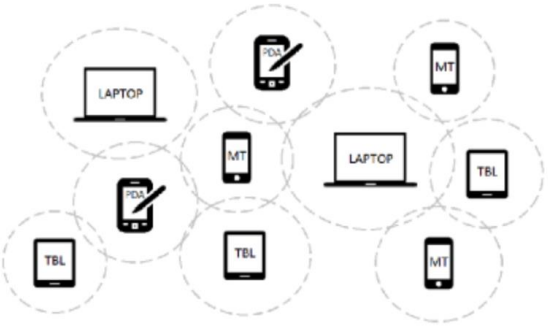


Fig. 1 Example of mobile ad hoc network

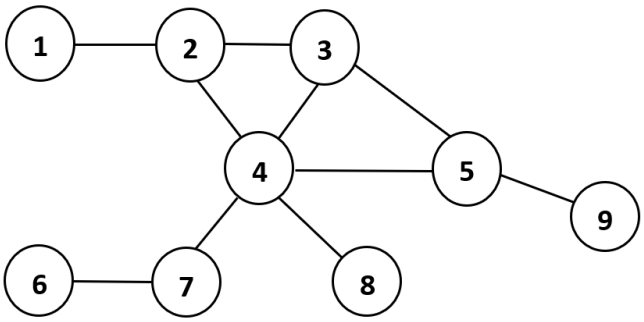


Fig. 2 Example of undirected graph of MANET network

Terminals in MANET networks are also referred to as nodes. In Fig. 1 and Fig. 2, we can see that there is indeed no fixed infrastructure in the network. Nodes can communicate with each other wirelessly if they are within range of each other. Nodes that are not in range of each other can communicate through other nodes, that is, the network is multi-hop.

In general, MANET networks can be defined:

- the nodes in the manet network are still in motion, working in a dynamic topology
- Multi-hop communication - nodes can communicate through other nodes
- All devices are equal and work in the so-called peer-to-peer mode. They act as routers and are responsible for managing communication and generating data on the network

- The nodes communicate via a predetermined same medium, radio waves, infrared connections, etc.
- Limited resources - due to mobility, nodes have limited energy or memory resources [1], [2], [3], [4], [5].

III. 5G NETWORKS

The 5G network represents a new generation of mobile networks that promises a thousand-fold increase in capacity, access for more than one hundred trillion devices with a data rate of around 10 Gb / s and a delay of up to 1 ms. To achieve these values, several key technologies need to be used. The required data rate can be obtained using massive MIMO (Multiple-Input and Multiple-Output) technology, using millimeter wavelengths and new waveforms. These technologies will also support the development of other new technologies, such as virtual reality, autonomous vehicles and autonomous drones or IoT. A radical increase in capacity and bandwidth will be achieved by merging multiple cellular systems and wireless networks into a single network.

Another important factor in the 5G network is spectrum efficiency. This parameter tells you how many bits per Hz per second can be transmitted over the air. Increasing this parameter also increases the capacity of the network without increasing the number of cells or increasing the width of the spectrum. Several aspects affect spectrum efficiency. It is e.g., modulation, multi-antenna communication, signal shape, inter-node coordination and spectrum control. The LTE (Long-Term Evolution) standard used already achieves high spectrum efficiency, but it still has certain reserves that the arrival of the 5G network should minimize.

Advantages of 5G networks:

- increased transfer rate
- high spectral efficiency of the system
- connecting more devices
- lower network delay
- better coverage of complex geographical areas
- lower energy consumption
- higher reliability of communication systems

These features and the main advantages of fifth generation networks are achieved mainly through the development of the above-mentioned MIMO technologies, millimeter waves and beam shaping. [6], [7], [19].

IV. EXAMPLES OF ALTERNATIVE TECHNIQUES FOR ROBUST DATA COMMUNICATIONS

A. Deep reinforcement learning (DRL)

Behavioral psychology has inspired a new machine learning technique, namely the strengthening of learning, which has proven to be an advanced technology and a new subcategory of machine learning. Fig. 3 shows the learning method of reinforcement learning. The DRL deals with the reward or benefit of agents associated with the environment through perceptions and actions that lead to adaptation. In reinforcement learning (RL), the agent seeks to optimize the long-term goal by interacting with the environment through a trial-and-error process.

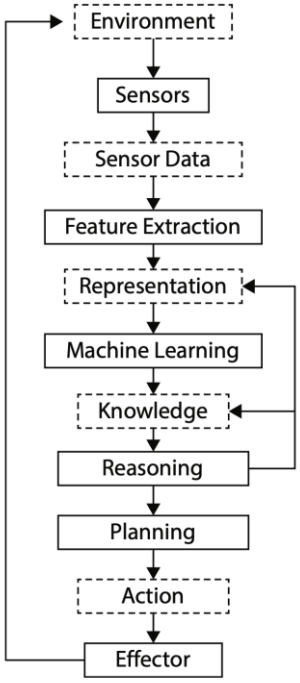


Fig. 3 Learning method of deep reinforcement learning [9]

B. Q-learning

One of the most commonly adopted reinforcement learning algorithms is Q-learning. Specifically, the RL agent interacts with the environment to determine the Q values on the basis of which the agent performs the action. The value of Q is defined as the discounted cumulative reward, starting from the n-tuple of the stock and the stock and then following a particular policy. When the Q values are learned after sufficient time, the agent can, in the current state, make a quick decision by performing the action with the largest Q value.

According to [14], the value of Q is updated as follows:

$$Q_{t+1}^{x}(s,a) = (1-\alpha)\, Q_t^x(s,a) + \alpha(R_a^x + \gamma\, max\, Q_t^a(s,b))$$

where $Q_{t+1}^{x}(s,a)$ is the value of Q for the current node, x selects the next α jump for the target s at $t+1$ on iteration, and R_α^x is the reward obtained, $maxQ_t^a(s,b)$ is the maximum possible future value for the next α jump. The scope of α learning indicates the importance of rewarding for previous implementations with the same policy.

C. Game theory

Game theory is an area of applied mathematics that deals with the study and analysis of situations whose interests can be defined as conflicting. More precisely, game theory focuses on the mathematical capture of the behavior of individuals in strategic situations, where the success and decision-making of one individual depends on the decisions and choices of other individuals.

In game theory, individuals (players) are characterized by the following characteristics:

- Each player is defined by one or more strategies

- There is a combination of moves (strategies) of all players in the game that leads to an optimal state (win, lose or draw)
- Each player has a certain profit from their optimal state
- All players behave rationally

Games can generally be divided into cooperative and non-cooperative. In a non-cooperative game, the player disregards the strategies of other players and behaves in such a way as to maximize his profit. In a cooperative game, on the other hand, players work together and look for a combination of their strategies so that everyone ultimately benefits from them. [8], [9], [10], [11], [12], [13].

D. Blockchain technology

Blockchain is the basic technology of the first cryptocurrency Bitcoin, and it is expected that applications such as IoT, Cloud and Edge Computing, Ad hoc networks, etc. will be based on blockchain in the future. It is divided into two types, a public blockchain to which free access is allowed, a user can view or contribute to its content, and a private blockchain, which can only be accessed with permission.

In a blockchain, transactions are wrapped in blocks. Each block contains the hash value of its previous block, which eventually results in a hash called a "block-chain". Fig. 4 provides a common blockchain structure.

Properties of blockchain:

- immutability
- distributed and trusted environment
- privacy and anonymity
- faster transactions
- reliable and accurate data
- transparency

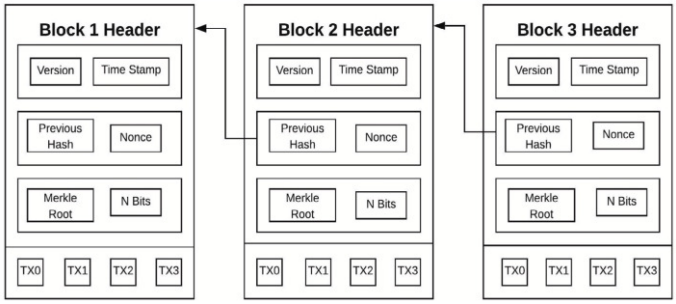


Fig. 4 Block data structure

The two best known types of consensus blockchain mechanism are Proof of Work (PoW) and Proof of Stake (PoS).

E. Proof of Work

Proof of Work is a mechanism by which miners solve a complex mathematical problem, which is represented by a block of transactions. The block of transactions is filled with unauthenticated transactions, and such a block reaches the miners, who use the hash function to transfer a large amount of data to the hash function. A hash is a short string of symbols. Creating a hash is simple, but in the proof of work system, the so-called nonce. Nonce is a short text that makes no sense but changes the hash results and thus the overall complexity. Such an algorithm can be based on the SHA 256 hash function respectively SHA 256d.

F. Proof of Stake

In the Proof of Stake system, new blocks are verified by the users themselves. The algorithm selects the user who created the block and inserts part of the held reward (cryptocurrency) as a backup and other user approve it. In this system, such a high computing power is not required as with PoW. The reward for the extracted block is not a new cryptocurrency, but fees for the completed transaction [15], [16], [17], [18], [19], [20], [21].

V. Network simulator NS3

For the purpose of obtaining the simulation of 5G networks together with implemented machine learning techniques and with a blockchain database we will use the NS3 program. This program is a network simulator with discrete events. NS3 is open-source project written in C++ programing language with Python bindings. This simulator uses modular kernel that is easily extensible using C++ or Python support. Main advantages are the use of the simulator in the creation and testing of the analytical part of the project. When teaching through NS-3, students can visually see the work of the network protocol and understand the effects of various environmental or other factors on the network and can also demonstrate the advantages and disadvantages of different strategies through comparison. On this platform, an experimenter can obtain an "ideal" network environment by configuring environment parameters and can monitor and record important information on the key node in real time to get first-hand information about evaluating network performance. The biggest disadvantage is that it does not have a graphical user interface. To analyze the results obtained, we must implement various, such as GNUPlot for plotting or FlowMonitor for obtaining parameters such as throughput, data rate, error rate, etc. [22], [23].

VI. Further research direction

In the next work, we would like to focus on alternative data transmission methods using different routing protocols for multi-hop networks based on machine learning, game theory and blockchain technology. In the main part of the work, we would like to simulate 5G networks with individual methods of data transmission using implemented techniques.

Acknowledgment

The research presented in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project VEGA 1/0753/20, KEGA 009TUKE-4/2019 and by Slovak Research and Development Agency project APVV-15-0517.

References

[1] J. He, S. Ji, Y. Pan, Y. Li, "Wireless Ad Hoc and Sensor Networks Management, Performance, and Applications" Taylor & Francis Group, New York, 2014.

[2] S. Bosagni, M. Conti, S. Giordano, I. Stojmenovic, "Mobile Ad Hoc Networking" second edition IEEE, New Jersey 2013, pp. 3-228.

[3] A. Pal, P. Dutta, A. Chakrabarti, J. P. Singh, “An efficient load balanced stable multi-path routing for mobile ad-hoc network” Springler-Verlag GmbH Germany, part of Springer Nature 2020.
[4] R. Gotzhein, “Real-time Communication Protocols for Multi-hop Ad-hoc Networks. Wireless Networking in Production and Control Systems” Technical University Kaiserslautern. ISBN 978-3-030-33319-5. 2020.
[5] M. M. A. Azim, X. Jiang, “Wireless Sensor Multimedia Networks, Architectures, Protocols, and Applications” Taylor & Francis Group, New York, 2016.
[6] V. K. Quy, N. T. Ban, N. D. Han, “An Advanced Energy Efficient and High Performance Routing Protocol for MANET in 5G” Journal of Communications Vol. 13, No. 12. 2018.
[7] M. A. Archana, T. Dineshkumar, V. Uma, “MANET for 5G Communication Networking” Department of Electronics and Communication Engineering, India. 2020.
[8] R. Dudukovich, CH. Papachristou, “Delay Tolerant Network Routing as a Machine Learning Classification Problem” Ohio, 2018.
[9] K. K. Singh, A. Singh, K. Cengiz, D. N. Le, “Machine Learning and Cognitive Computing for Mobile Communications and Wireless Networks”. 2020, pp. 1-66.
[10] J. Jagammath, N. Polosky, A. Jagannath, F. Restuccia, T. Melodia, “Machine learning for wireless communications in the Internet of Things: A comprehensive survey”, Elsevier, 2019.
[11] T. K. Rodrigues, K. Suto, H. Nishiyama, J. Liu, N. Kato, “Machine Learning meets Computation and Communication Control in Evolving Edge and Cloud: Challenges and Future Perspective”, IEEE accepted for publication, 2019.
[12] S. Zeadally, M. Tsikerdekis, “Securing Internet of Things (IoT) with machine learning”, College of Communication and Information, University of Kentucky, 2019.
[13] S. N. Hasnain, “A Survey on The Application of Machine Learning in Ad hoc Networks Routing”, Electical Engineering and Information Technology, Technical University Ilmenau, 2014.
[14] K. Wang, W. Ch. Wong, T. Y. Chai, "A MANET Routing Protocol using Q-Learning Method Integrated with Bayesian Network", IEEE International Conference on Communication Systems (ICCS), 2012, pp. 270-274.
[15] K. Asaf, R. A. Rehman, B. S. Kim, “Blockchain technology in Named Data Networks: A detailed survey”, Journal of Network and Computer Applications, 2020.
[16] C. Machado, C. M. Westphall, “Blockchain incentivized data forwarding in MANETs: Strategies and challenges”, Elsevier, 2020.
[17] A. Kumari, R. Gupta, S. Tanwar, N. Kumar, “A taxanomy of blockchain-enabled softwarization for secure UAV network”, Elsavier, 2020.
[18] M. T. Lwin, Y. B. Ko, D. Kim, “When Blockchain Takes Care of the OLSR Network”, Department of Computer Engineering Ajou University. Department of Computer Science Kangwon National University, South Korea, 2019.
[19] P. Mehta, R. Gupta, S. Tanwar, “Blockchain envisioned UAV networks: Challenges, solutions, and comparisons”, Elsevier, 2020.
[20] A. Sajid, B. Khalid, M. Ali, S. Mumataz, U. Masud, F. Qamar, “Securing Cognitive Radio Networks using blockchains”, Elsevier, 2020.
[21] M. Li, F. R. Yu, P. Si, W. Wu, Y. Zhang, “Resource Optimization for Delay-Tolerant Data in Blockchain-Enabled IoT with Edge Computing: A Deep Reinforcement Learning Approach”, University of Exeter, IEEE, 2020.
[22] L. Campanile, M. Gribaudo, M. Iacono, F. Marulli, M. Mastroianni, “Computer Network Simulation with ns-3: A Systematic Literature Review”,Electronics 2020.
[23] X. Rui, “Performance Analysis of Mobile Ad Hoc Network Routing Protocols Using ns-3 Simulations”, Electrical Engineering & Computer Science, University of Kansas, 2019, pp. 6-59.

Overview: Text Classification using Deep Learning Models

[1]*Viera Maslej Krešňáková (3rd year),*
Supervisor: [2]*Peter Butka*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]viera.maslej.kresnakova@tuke.sk, [2]peter.butka@tuke.sk

***Abstract*—The emergence of anti-social behavior in online environments presents a serious issue in today's society. This paper presents an overview of publications focused on the detection and classification of antisocial behavior and emotion analysis. Introduced publications deal with the use of deep learning methods for text processing. As part of solving the problem in the individual papers, we created several different approaches, compared, e.g., the use of preprocessed and unprocessed text and various types of neural networks.**

***Keywords*—deep learning, text-processing, antisocial behavior, fake news, toxic comments, Slovak online space**

I. Introduction

Neural networks are considered to be one of the best-performing machine learning algorithms. They have brought great success in the field of artificial intelligence, such as in the field of computer vision, where their task is image processing and pattern recognition, and, for example, in sound processing and speech recognition. In our works, we took a closer look at how neural networks can be used to work with the textual data.

In this paper, I present an overview of my published papers in the text-processing domain. Most of them deal with the detection and classification of antisocial behavior in the online space. In conference papers [1] and [2], we focused on the Slovak online space. We continue to develop these experiments and, at the same time, work on obtaining a larger number of quality annotations of just Slovak texts. In addition to political and economic news Slovak online space, we also focus on detecting fake news in articles focused on topics around Covid-19.

This paper presents two publications from the Current Contents Connect database [3], [4], a book chapter [5], and CINTI [6], SAMI [1], and ICETA [2] conference papers.

II. Achieved results

A. Addressing false information and abusive language in digital space using intelligent approaches

Nowadays, more whenever in the past, important to deal with various unhealthy phenomena, that prohibit effective communication and knowledge sharing in the digital space. Undesired user behaviour and user-generated content in the online environment (mostly on social media) can have various forms, probably, the most harmful is the creation and spreading of false information (e.g., fake news) and using abusive language (e.g. hate speech). In book chapter [5], we provided a summary of existing research works, challenges and open problems. We introduce our research results addressing false information and abusive language. Our approaches are based on intelligent and knowledge-based methods, mainly machine learning, natural language processing, and semi-automatic approaches.

B. Comparison of deep learning models and various text pre-processing techniques for the toxic comments classification

In work presented in paper [3], we aimed to compare and evaluate different current state-of-the-art models for multi-label toxic comments classification. We experimentally evaluated the performance of deep learning models, including composed architectures with different methods of text representation and pre-processing. On top of that, currently, popular transformer language models, such as BERT [7] and its modifications were compared as well. We aimed to explore the assumption that in tasks such as detection of anti-social behaviour in the online environments, the application of traditional pre-processing techniques could lead to loss of particular specific information characteristic for such behaviour. We aimed to explore the influence of different pre-processing and representation methods on the deep learning and transformer models also in multi-label task aimed to detect the specific type of anti-social behaviour (in this case, toxic comments). We experimentally evaluated composed architecture of BiLSTM+CNN network with different text representations, pre-trained embeddings and compared it with BERT and its variants.

C. Deep learning methods for fake news detection

The work presented in paper [6] is aimed to use deep learning techniques to tackle the problem of the detection of fake news from the text. We trained different neural network models (feedforward, convolutional, and LSTM [8]) on data containing the full text of the analyzed articles as well as only title texts. The models were trained using a labeled dataset of fake and real news, and such models proved to be effective in this task. When comparing the evaluation metrics, most of the models gained consistent performance; however, convolutional and LSTM models proved to be the most effective. When comparing the evaluation metrics on the full-text data to only title texts, the models still managed to perform on a similar level. On the other hand, the effect of using just the title texts for training proved to be effective during the training phase. This could be significant in real-world tasks, when using much

larger training data or when the deployed models have to be updated frequently to adjust to incoming data.

D. Annotated dataset for the fake news classification in Slovak language

Paper [2] presented the collection of the data and creation of the labelled dataset for the fake news detection in Slovak language using crowdsourcing approach. We have utilized the Doccano crowdsourcing platform to obtain the annotations for 1535 articles. The annotations from 20 human annotators were summarized and we used the labelled dataset to train the neural network classifier during a set of preliminary experiments. Labelling of the data using the crowdsourcing enables us to create the datasets for such tasks, however, some issues have to be taken into the consideration. At first, the proper selection of the data (articles) is crucial. In this work, we downloaded multiple articles, which could not be considered as the news articles (e.g. blogs), which had to be excluded from the dataset. It is also necessary to involve a larger group of annotators to obtain as much annotations for the particular articles as possible to remove the bias, which may be present in small group of annotators.

E. Fake news detection in Slovak language using deep learning techniques

The work presented in paper [1] described the use of deep learning models for detection of the fake news from the Slovak news articles. We approached this problem as a binary classification task, by predicting if a news piece represents the regular news article or if it contains false/misleading information. In order to train such models, we needed to create the dataset consisting of both types of news, which we obtained from the Slovak online news portals. Very important is the labelling of the samples. Many datasets are created by human annotation, which may lead to bias, incorporated by the annotators themselves. We used a combined approach, in which we labelled the news according to their publisher in the first phase and then tuned the assigned labels manually by a more in-depth exploration of the content of the article. Then, we trained two deep learning architectures using this dataset - CNN [9] and LSTM neural networks [8]. We used only the text content of the articles and applied standard pre-processing methods, which included tokenization, lowercasing and stopwords removal. The models were validated and evaluated using train/validation/test split, standard classification metrics and confusion matrix were used. LSTM architecture achieved superior performance, managing to detect most of the false articles while producing less false negatives as the CNN approach.

F. Emotion analysis in human-robot interaction

Paper [4] connects two large research areas, namely sentiment analysis and human–robot interaction. Emotion analysis, as a subfield of sentiment analysis, explores text data and, based on the characteristics of the text and generally known emotional models, evaluates what emotion is presented in it. There are several approaches and algorithms to detect emotions in the text data. We decided to apply a combined method of dictionary approach with machine learning algorithms. As a result of the ambiguity and subjectivity of labeling emotions, it was possible to assign more than one emotion to a sentence; thus, we were dealing with a multi-label problem. Based on the overview of the problem, we performed experiments with the Naive Bayes, Support Vector Machine and Neural Network classifiers. Results obtained from classification were subsequently used in human–robot experiments.

III. Conclusion

My dissertation focuses on deep learning methods in analytical tasks and decision making. One of the areas I focus on is natural language processing – text processing. A great benefit for my work is publication [3], where I focus on comparing popular transformers models such as BERT and traditional bidirectional Long-Short-Term Memory networks in combination with one-dimensional convolutional neural networks. We also compared different methods of pre-processing. I applied the acquired knowledge in other publications.

My work continues by uncovering antisocial behavior in the Slovak online space. I believe that the result will be beneficial, especially in detecting fake news on topics related to the Covid-19 pandemic.

Acknowledgment

This work was supported by Slovak Research and Development Agency under the contract No. APVV-16-0213 and by the Slovak VEGA project under grant No. 1/0685/21.

References

[1] K. Ivancová, V. Maslej-Krešňáková, and M. Sarnovský, "Fake news detection in slovak language using deep learning techniques," in *SAMI 2021: IEEE 19th World Symposium on Applied Machine Intelligence and Informatics*, 2021.

[2] M. Sarnovský, V. Maslej-Krešňáková, and N. Hrabovská, "Annotated dataset for the fake news classification inslovak language," in *ICETA 2020: 18th International Conference on Emerging eLearning Technologies and Applications*, 2020.

[3] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, "Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification," *Applied Sciences*, vol. 10, no. 23, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/23/8631

[4] M. Szabóová, M. Sarnovský, V. Maslej-Krešňáková, and K. Machová, "Emotion Analysis in HumanâĂŞRobot Interaction," *Electronics*, vol. 9, no. 11, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/11/1761

[5] K. Machova, I. Srba, M. Sarnovský, J. Paralič, V. Maslej-Kresnakova, A. Hrckova, M. Kompan, M. Simko, R. Blaho, D. Chuda *et al.*, "Addressing false information and abusive language in digital space using intelligent approaches," in *World Symposium on Digital Intelligence for Systems and Machines*. Springer, 2020, pp. 3–32.

[6] V. Maslej Kresnakova, M. Sarnovsky, and P. Butka, "Deep learning methods for fake news detection," *2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo)*, pp. 143–148, 2019.

[7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, no. 8, pp. 1735–1780, 1997.

[9] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.

Improving Program Comprehension Using Testing Methods and Tools

[1]*Filip GURBÁL' (1st year),*
Supervisor: [2]*Jaroslav PORUBÄN*

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

[1]filip.gurbal@tuke.sk, [2]jaroslav.poruban@tuke.sk

***Abstract*—Computer programming is very demanding activity for comprehension, which makes it difficult for programmers to develop new software and also maintain it. This paper looks for the methods and tools that can simplify program comprehension. Since we believe, that testing can provide useful information to lower the cognitive load, we describe what causes high cognitive demand and analyse researches that explores program comprehension from perspective of testing. At the end of this paper we describe current known methods and tools and suggest next steps in this research.**

***Keywords*—cognitive load, IDE, program comprehension, quality assurance, testing**

I. Introduction

Complexity of computer programming has greatly evolved due to the increasing complexity of software systems. There are studies focused on support for software development, bringing methods to improve processes, programming languages, programming tools etc (such as [1]). As a result of growing complexity in software, a new area called *program comprehension* has emerged. The programmer during development tries to meet customer's requirements by writing program code, which is only an abstract description of the solution (semantics). This brings a gap between problem and solution domain which programmers must solve [2].

During the development, a programmer may need to change code after previous programmer and this especially occurs during software maintenance. In this case, programmer must additionally understand the existing code of the existing solution. According to the latest neuroscientific measurement [3], this process of understanding is very cognitively demanding task. The aim of researching program comprehension is to find ways to reduce cognitive load and improve software understanding, thereby accelerating software development process and lower the cost.

In Section II we will describe *program comprehension* and how it relates to *reverse engeneering*. In Section III we mention some frequently asked questions of programmers during program comprehension process. Section IV describes testing methods and what information they can provide to improve program comprehension. In Section V we present researches dealing with the use of tests to improve program comprehension. Finally in Section VI we outline the future direction of this research.

II. Program comprehension

Based on the specification, the programmer creates a mental model of the program in his/her head with program-specific properties and problem domain knowledge [4]. Using this model, the programmer creates a software artifact in the form of program code. The code is then compiled into an executable application that tries to meet requirements defined in the specification. If the application does not behave as required, or the specification is extended, it is necessary to modify or extend the existing program code. For modifying an existing code created by someone else the programmer must first comprehend what it does. Program comprehension is the state when the programmer understands the structure, behavior and connection to the application domain [5]. When a programmer creates or modifies a solution to a problem in the application domain, it is important for him/her to understand its critical aspects and the possible impact of the code on different parts of the application. Therefore understanding application domain is also the key to a successful software development [6]. However, it is not necessary to know entire code of the program.

Several studies pointed out that the process of program comprehension takes up to a half of the programmer's time during their work with the code [1], [7]. According to Blackwell [8], one of the main causes of high cognitive demand in programming is the use of notation. Since the program code is only an abstract description of the action (as we stated before), the result of this action (and state of the program after its execution) is not directly visible to the programmer during development. This means, that one of the key features of a programmer to create a (good) software system is the ability to create program abstractions [9]. This cognitive load can be basically reduced by keeping the program code readable from the start of the development [10], [11].

In connection to the software maintenance there is another area called *reverse engineering*. This area deals with the analysis of the existing software systems in order to collect metadata and create tools for visualisation and interpretation of various program aspects [12]. It is closely related to program comprehension, reverse engineering provides tools which helps with comprehension and at the same time program comprehension analyzes effectiveness of these tools to improve them and add additional knowledge of development practices.

III. Frequently asked questions of programmers

When trying to comprehend program, programmers often ask questions in their heads about the program code. Many studies analyze the nature of these questions so that we can better focus on what is important for programmers. Sillito et al. [13] constructed a list of 44 basic questions. They are related to source code (e.g., "Does this type have any siblings in the type hierarchy?"), run-time properties of the program ("What are the values of these arguments at runtime?"), and their interconnection ("Where in the code is the text of this error message?"). The authors found that the majority of these questions have only partial support, and almost every question requires a different tool. The use of large number of different tools causes unnecessary cognitive overhead.

LaToza and Myers [14] found that programmers often ask about reachability, which means that they try to find an expression in the process of program execution. They also collected 94 hard-to-answer questions [15]. These are related to implementation, debugging, rationale behind the code, refactoring, testing, compilation, performance etc. The frequency of these questions can is useful for prioritization when designing supporting tools.

IV. Quality assurance in software

Important part of the software development life cycle is testing. The goal of testing is quality assurance – ensuring reliability, correctness, validation and verification [16]. Testing process is analysis of the software from different perspectives with aim to check its correct behavior. If software behavior does not match the specification, we speak of a software error [17]. Programmer modifying existing program code may undesirably affect other parts of the system causing different (wrong) behavior. If tests are included in the program, they can detect errors and notify the programmer about impact of his/her modifications during the development process.

The quality of software is closely related to the program comprehension. Tools provided by integrated development environment (IDE), programmer's experience, programming language and program understanding have a direct or indirect impact on the quality of product. Tests contains lot of information about tested software, so we are trying to use them for improving program comprehension.

Testing (analysis of software) is divided into *static* and *dynamic*. Static testing is performed on a non-running program code and it checks the status of the written code. This testing includes reviews, inspections and guidelines. The opposite of static testing is dynamic testing, which performs analysis on the running program and monitors program execution and result. Lewis [18] in his book defines and analyzes 60 testing techniques. The most popular are testing techniques are black box, white box, grey box [19], regression [20], usability, performance, unit etc. The individual techniques can be combined, what provides commonly used main types of tests like functional testing, performance testing or security testing [21].

We will focus on functional testing, because we assume that this type of testing most likely contains useful information about the system under test within its implementation, as we explain later in Subsection IV-A. However, we do not rule out the possibility that other types of testing may also be useful for improving program comprehension, but since they focus on different system features, such as execution time, system load, environment performance, security etc., we do not expect much benefit from them.

A. Functional Testing

Functional testing verifies behavior and functionality of the application. Tests are created based on customer's specification. Main subtypes of functional testing are:

1) *Unit Testing*. This type of tests analyzes the smallest unit of program code independently of the rest of the program (small part – units – are isolated).
2) *Integration Testing*. This type of tests verifies communication and dataflow between components.
3) *System Testing*. This type of tests validates the complete and fully integrated software product.
4) *Acceptance Testing*. This tests checks whether required functionality can be performed by the customer.

We can apply to almost any type of testing white box, black box or gray box testing technique [19]. We can test software from the perspectives that "how" certain actions are performed, or "what" is the result of the actions. White box testing is used to verify the internal structure of the code and its internal logic. Author of white box tests mus be familiar with the code and the technical specifications. This testing checks "how" the software is supposed to behave.

The black box testing is used to easily verify the correctness of the code. It simply compares the output of the action with the expectation regardless of program's internal processes. This testing has no knowledge about "how" the action is performed. This makes these tests easier to create and understand, since author or reader does not need to know the code or the technical specifications. In this case, also a non-programmer can be a tester.

The combination of these two testings is gray box testing [19]. It verifies the output of the performed action with limited knowledge of internal processes.

Each type of testing is closely linked to the program code, which gives us more information about software. Functional tests basically describe the problem, so they can be used as a specification or documentation. Also, thanks to the fact that black box testing can be created by a non-programmers, tests can provide information about software from non-programmer perspective, which can increase program comprehension process.

V. State of the art

The topic of using tests to improve program comprehension has been the subject of some researches in the past, but it is still not widely researched area. Benedusi et al. in his paper [22] investigates the role of testing and dynamic analysis in the process of program comprehension. The subject was explored in the context of their Esprit DOCKET Project (DOcument and Code Knowledge Elicitation Toolset), whose aim was to design and experiment second-generation reverse engineering processes combining multiple knowledge sources. In their paper they stated four basic observations:

1) Test suites should be available at maintenance time at least for regression testing purposes,
2) Maintenance programmers regularly use program execution to support program comprehension, even when

documents and various reverse engineering tools are available. Natural language documents are very helpful in program comprehension, but they may be out of date or inaccurate.

3) Test cases are the passive or active subject of existing software life cycle phases (specification, test case design, debugging and error analysis, change request formulation, change analysis, regression testing) producing various kinds of peculiar and valuable knowledge.
4) Test cases are important starting point for the capture and reuse of knowledge acquired empirically and lessons learned during the operational life of the software.

In the DOCKET method they used test cases and dynamic analysis as necessary and convenient support for program comprehension, because they allowed highlighting relationships between elements in the software, which are difficult to obtain by static analysis, provided direct acquisition of functional knowledge, helped map functions with code structures or vice versa, provided examples effective in the comprehension and learning process etc. They researched these subjects from two perspectives:

- *The improvement of the reverse engineering processes* – according to the paper, test cases can reveal relationships that are only visible during program execution, but on the other hand, tests can track only a subset of feasible program execution paths.
- *Reusing the knowledge gained during software development* – authors in the paper argue that test developers during software development have to understand the system well to create good tests and define what is and what is not an error. Their knowledge can be reused to improve program comprehension process.

They concluded that the use of well-designed regression tests can be useful for improving program comprehension. Another paper from Sneed [23] confirms, that test developer must have broad knowledge about the tested system. He researched the program comprehension for the purpose of testing and pointed out that in order for the tester to achieve the best possible coverage of the program code, he must find as many ways as possible to detect different errors. The tester needs to gather a large amount of information in order to answer the questions needed for test development (what are the objects that need to be tested, how to test them optimally and effectively, to what degree must each object be tested). The author focuses on the specific requirements for the tester in terms of testing techniques (white box and black box testing). The paper describes that each team member focuses on their part of the software (analysts focus on specification, programmers on their part of code and interfaces), but only testers need to have a much broader knowledge of the system. This implies that tests contain lot of information about system and can be used as a documentation.

In the book Software evolution [24], Moonen et al. write about software evolution and the impact of testing phase on program comprehension. He looks at testing in an agile environment (where extreme programming method is in use) and listed a few benefits of automated testing in connection with program comprehension:

- testing policy encourages programmers to explain their code using test cases, instead of writing comments and updating documentations,

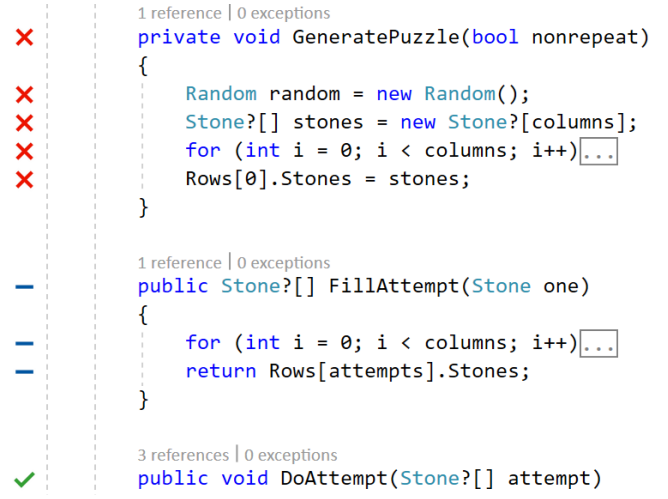


Fig. 1. Live Unit Testing feature in IDE Visual Studio 2017 Enterprise.

- the requirement that all tests must run 100% at all times, ensures that the documentation via unit tests is kept up-to-date,
- adding unit tests provides a repeatable program comprehension strategy, because programmer can experiment their hypotheses about code understanding, and
- a comprehensive set of unit tests reduces the comprehension space when modifying source code, programmers are more confident when modifying code, because they know that tests will find new errors.

In the book they also explained the importance of code refactoring, which is modifying the code without changing the functionality in order to improve quality of the software. Such modifications effectively increase readability of the code, which reduces cognitive load (as said in Section II). Authors points out that refactoring may also adversely change the behavior of the program and introduces "Test-Driven Refactoring", which tells that it is better to first modify the structure of tests and then modify the production code (similar approach as in Test-Driven Development method, where tests are developed first and based on them is developed program code).

All these researches show that tests and quality assurance are very useful in program comprehension. Our goal is to obtain information from tests and link them into the program code inside the integrated development environment (IDE). The IDE is an environment that helps programmers with development by providing useful tools, such as syntax highlighting, class explorer, search for references ("where is this used?"), automated refactoring, debugging and so on. There are already small tools which links tests with the source code integrated within some IDEs. For example, the Live Unit Testing feature in the Enterprise version of Visual Studio 2017 shows the test results directly in the source code along with information about the coverage (Figure 1).

VI. Future Directions

There are many tools that help with program comprehension, but as mentioned in Section III, separate tools are usually used to answer different questions. In addition, we did not find any research focused on linking tests to the main source code. We must first explore existing tools and methods to improve program comprehension.

A. Code Projections

One of interesting solutions is filtering source code with concern-oriented projections [25]. Projections allow programmer to collect various parts of code from different source code files based on his/her concern. With this approach we can filter those part of code that are relevant to tests and connect them.

B. Finding tests in code

First issue for this research is find tests within project codes. Our department addressed this topic in papers [26], [27], [28] and analyzed the presence of tests in projects using various natural language processing (NLP) methods.

C. Next steps

Next step is to explore testing tools and their usage. Our focus will be to find possibly the most universal tool or method for detecting presence of tests in various projects, then we will be able to analyze how are these tests implemented and what information they can provide. We believe that tests reflect real problem domain scenarios, so it should be possible to gather useful information about system. Our main goal is to find the tool or method that will interpret these information included in tests to improve program comprehension.

Acknowledgement

This work was supported by project VEGA No. 1/0762/19 Interactive pattern-driven language development.

References

[1] T. Kosar, M. Mernik, and J. C. Carver, "The impact of tools supported in integrated-development environments on program comprehension," in *Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces*, 2011, pp. 603–608.

[2] J. Greenfield and K. Short, "Software factories: Assembling applications with patterns, models, frameworks and tools," in *Companion of the 18th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications*, ser. OOPSLA '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 16–27. [Online]. Available: https://doi.org/10.1145/949344.949348

[3] N. Peitek, J. Siegmund, S. Apel, C. Kästner, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann, "A look into programmers' heads," *IEEE Transactions on Software Engineering*, vol. 46, no. 4, pp. 442–462, 2020.

[4] R. W. Collins, A. R. Hevner, G. H. Walton, and R. C. Linger, "The impacts of function extraction technology on program comprehension: A controlled experiment," *Information and Software Technology*, vol. 50, no. 11, pp. 1165–1179, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584908000530

[5] T. J. Biggerstaff, B. G. Mitbander, and D. Webster, "The concept assignment problem in program understanding," in *Proceedings of 1993 15th International Conference on Software Engineering*, 1993, pp. 482–498.

[6] R. Offen, "Domain understanding is the key to successful system development," *Requirements Engineering*, vol. 7, no. 3, pp. 172–175, 2002.

[7] R. Tiarks, "What maintenance programmers really do: An observational study," 2011.

[8] A. F. Blackwell, "What is programming?" in *In Proceedings of PPIG 2002*. MIT Press, 2002, pp. 204–218.

[9] J. Kramer, "Is abstraction the key to computing?" *Commun. ACM*, vol. 50, pp. 36–42, 04 2007.

[10] D. Roy, S. Fakhoury, J. Lee, and V. Arnaoudova, "A model to detect readability improvements in incremental changes," in *Proceedings of the 28th International Conference on Program Comprehension*, ser. ICPC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 25–36. [Online]. Available: https://doi.org/10.1145/3387904.3389255

[11] J. Johnson, S. Lubo, N. Yedla, J. Aponte, and B. Sharif, "An empirical study assessing source code readability in comprehension," 09 2019, pp. 513–523.

[12] H. Müller and H. Kienle, "A small primer on software reverse engineering," *Technical Report, University of Victoria*, 01 2009.

[13] J. Sillito, G. C. Murphy, and K. De Volder, "Asking and answering questions during a programming change task," *IEEE Transactions on Software Engineering*, vol. 34, no. 4, pp. 434–451, 2008.

[14] T. D. LaToza and B. A. Myers, *Developers Ask Reachability Questions*. New York, NY, USA: Association for Computing Machinery, 2010, p. 185–194. [Online]. Available: https://doi.org/10.1145/1806799.1806829

[15] ——, "Hard-to-answer questions about code," in *Evaluation and Usability of Programming Languages and Tools*, ser. PLATEAU '10. New York, NY, USA: Association for Computing Machinery, 2010. [Online]. Available: https://doi.org/10.1145/1937117.1937125

[16] M. Ehmer and F. Khan, "A comparative study of white box, black box and grey box testing techniques," *International Journal of Advanced Computer Science and Applications*, vol. 3, 06 2012.

[17] R. Patton, *Software Testing*. USA: Sams, 2000.

[18] W. E. Lewis, *Software Testing and Continuous Quality Improvement, Third Edition*, 2nd ed. USA: Auerbach Publications, 2008.

[19] M. Khan, "Different forms of software testing techniques for finding errors," *International Journal of Computer Science Issues*, vol. 7, 05 2010.

[20] A. Sawant, P. Bari, and P. Chawan, "Software testing techniques and strategies," *International Journal of Engineering Research and Applications(IJERA)*, vol. 2, pp. 980–986, 06 2012.

[21] I. Hooda and R. S. Chhillar, "Software test process, testing types and techniques," *International Journal of Computer Applications*, vol. 111, no. 13, 2015.

[22] P. Benedusi, V. Benvenuto, and L. Tomacelli, "The role of testing and dynamic analysis in program comprehension supports," in *[1993] IEEE Second Workshop on Program Comprehension*, 1993, pp. 149–158.

[23] H. M. Sneed, "Program comprehension for the purpose of testing," in *Proceedings. 12th IEEE International Workshop on Program Comprehension, 2004.*, 2004, pp. 162–171.

[24] L. Moonen, A. Deursen, A. Zaidman, and M. Bruntink, "On the interplay between software testing and evolution and its effect on program comprehension," in *Software Evolution*, 2008.

[25] J. Porubän and M. Nosál, "Leveraging program comprehension with concern-oriented source code projections," *OpenAccess Series in Informatics*, vol. 38, pp. 35–50, 01 2014.

[26] M. Madeja and J. Porubän, "Tracing Naming Semantics in Unit Tests of Popular Github Android Projects," in *8th Symposium on Languages, Applications and Technologies (SLATE 2019)*, ser. OpenAccess Series in Informatics (OASIcs), R. Rodrigues, J. Janousek, L. Ferreira, L. Coheur, F. Batista, and H. G. Oliveira, Eds., vol. 74. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, pp. 3:1–3:13. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2019/10870

[27] M. Madeja and J. Porubän, "Accuracy of unit under test identification using latent semantic analysis and latent dirichlet allocation," in *2019 IEEE 15th International Scientific Conference on Informatics*, 2019, pp. 000 161–000 166.

[28] M. Madeja and J. Porubän, "Unit under test identification using natural language processing techniques," *Open Computer Science*, vol. 11, no. 1, pp. 22–32, 2021. [Online]. Available: https://doi.org/10.1515/comp-2020-0150

Application of Transformer neural architecture in speech recognition of Slovak language

[1]*Anton BUDAY (2nd year),*
Supervisor: [2]*Anton ČIŽMÁR*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]anton.buday@tuke.sk, [2]anton.cizmar@tuke.sk

***Abstract*—This paper discusses the prototype of an E2E (end-to-end) offline Transformer neural network architecture for Slovak language without a language model (LM). Transition to E2E ASR (automatic speech recognition), experimental results and next work proposal are described. Despite the small amount of training data, Transformer has achieved quite promising results comparing to conventional DNN/HMMs approach.**

***Keywords*—attention, deep neural networks, end-to-end, slovak language, speech recognition, transformer**

TABLE I
SUMMARIZING TABLE WITH RESULTS OF E2E TRANSFORMER PROTOTYPE WITHOUT EXTERNAL LM INTEGRATION

Parameter	Testing Set [%]	Validation Set [%]
CER	17.8	11.5
WER	50.9	39.9

This table illustrates the results of metrics CER & WER, respectively. CER stands for character error rate, WER stands for word error rate. Possible external LM integration should be realized in near future.

I. Introduction

One of the applications of machine learning technologies is speech recognition. According to J. Pierce from Bell Telephone Labs, a technology of ASR (automatic speech recognition) is attractive to money [1]. He designated it being similar to the attraction of schemes for turning water into gasoline, extracting gold from the sea, curing cancer, or even going to the Moon. This assertion is supported by a number of scientific papers being published and conferences being organized each year, such as INTERSPEECH, SPECOM, or ICASSP.

In the last decade, the development of ML (machine learning) technologies has grown drastically, therefore affecting lives of all of us. Older ASR approaches relied on conventional HMM/GMM statistical models. Since the advent of deep learning capable hardware, hybrid deep neural network (DNN)/GMM approaches have emerged. Recently in 2015, the novel approach to ASR was proposed, known as an E2E (end-to-end) ASR. The advantage of E2E ASR is that it allows for joint optimization of ASR system components, even, no LM is needed. With E2E ASR, it is also easier to embed it into mobile devices because of reduced data storage requirements. However, there are many disadvantages of E2E ASR. It is hugely impacted by the lack of training data. This is the main reason why conventional (hand-crafted) ASR is still on par with E2E speech recognition.

II. Research task status

Our team at Laboratory of Speech and Mobile Technologies from KEMT (located at FEI TUKE in Košice) have previously done an outstanding work on designing and maintaining offline and online hybrid DNN/HMM ASR models.

As a result, many speech and linguistic datasets are available. However, these have not been well-tried on novel, purely E2E ASR neural architectures. Hence, we have tried to train the prototype of an E2E-based ASR system on Slovak speech and linguistic dataset.

III. Offline Transformer for inference of Slovak spoken language

This section describes an experimental prototype of Transformer-based ASR system trained on Slovak dataset without any external language decoding (scoring).

A. Design procedure of Transformer ASR

Transformer is implemented in Pytorch framework embedded inside ESPnet. ESPnet follows so-called multitask learning (MTL) principle of ML.

The workflow of ESPnet to create ASR model is as follows:

1) **Resource data acquisition**. Data resources come from TUKE-BNews-SK corpus developed at KEMT. Dataset comprises acoustic data from prime time broadcast and sports news, social programs and TV shows. Original data are multichannel and recorded with 48 kHz sampling rate, but conventional model [2] required 16 kHz and primarily single-channel speech signals. Consequently, original waveform signals had to be downsampled. Recordings took place in various (noisy) environments, so principle of robustness and data sparsity is preserved. Paper [2] indicates conventional 21.94% WER. Apart from speech data, primarily language (text) data were formatted in Kaldi style (including speech waveform links). Utilized speech data have 256 kbit/s bitrate and are encoded with 16-bit signed PCM.

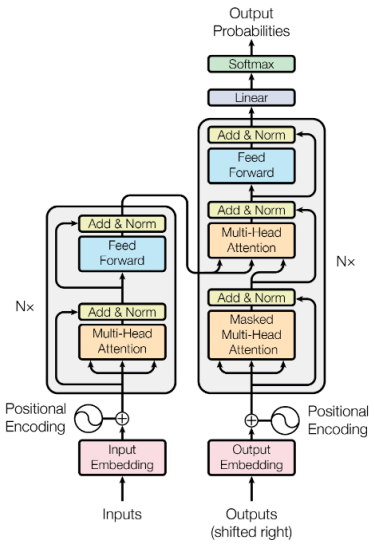


Fig. 1. Original implementation of Transformer in NLP applications by Google Brain engineers [4]. It contains encoder and decoder parts, multi-head attention [5], relative positional encoding, normalization, etc.

2) **Data formatting**. Formatting of data was done with Kaldi procedures, whereas data are divided intro three groups, that is: training, validation and test set. Part of training data was used as a validation set, because original dataset was divided only into two sets, namely training and test set.
3) **Speech features and normalization**. 40 bins of FBANK features were used as a representation for speech along with extracted pitch characteristics. Very long or short utterances were omitted. Cepstral features were normalized using CMVN normalization without any specific parameters.
4) **Training and decoding**. Training utilized ensemble learning methods (checkpoint ensembles), which improved performance of Transformer. Specifically, checkpoint averaging method was used [3]. Conformer network training was unsuccessful possibly due to unspecified bugs in ESPnet's code (or some incompatiblity with Slovak data). ESPnet depends on Pytorch deep learning framework, which contains an implementation of Transformer architecture. In spite of the fact that Transformers generally allow for parallel processing, GPU-based training lasted 10 days, because older NVIDIA K40 with 12GB VRAM was used. No LM was necessary during decoding (rescoring).

B. Results of Transformer-based end-to-end ASR

Fig. 1 describes implemented Transformer architecture. Similarly, Fig. 2 depicts performance of its training on validation dataset. The most important results are shown in Table I. Language-specific prototype of E2E Transformer achieved promising results. It is supported by comparison with conventional ASR system trained on the same data with LM, which achieved roughly 20% WER.

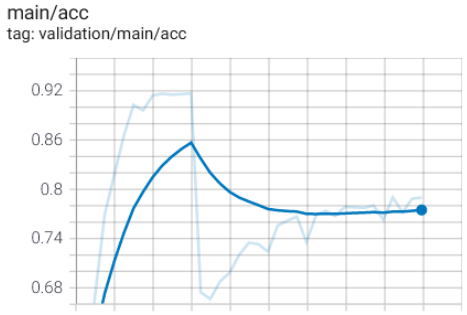


Fig. 2. Accuracy of Transformer ASR inference on validation dataset. Visually, it has achieved roughly 78% accuracy.

Correct	id: (jarka|vilhanova-jarka|vilhanova_bn-ba1-exp-spra-2010-05-12-xx-00-00-s0068) Scores: (#C #S #D #I) 28 0 0 0 REF: j a r k a <space>v i l h a n o v á <space>r á d i o <space>e x p r e s HYP: j a r k a <space>v i l h a n o v á <space>r á d i o <space>e x p r e s Eval:
Incorrect (substitution & deletion)	id: (jan|petrovic-jan|petrovic_bn-ke1-st1-spra-07-03-03-19-30-00-s0121) Scores: (#C #S #D #I) 2 4 2 0 REF: je TOUTO CHEMIKáLIOU AJ ZNEčISTENá pôda V AREáLI HYP: je ***** TO CHEMIKALIOVAJ ZNEčISTENA pôda * VAREáLI Eval: D S S S D S

Fig. 3. An example of accurate and inaccurate transcription inferred from Transformer E2E ASR without LM. Total amount of speech data involves 56 hours for training and 12 hours for testing datasets.

C. Proposal of next work

On the basis of this work we have set these dissertation goals to explore:

- usage of unlabelled automatically transcribed Slovak language data for novel E2E system training,
- addition of external LM in order to verify system accuracy alteration [6] [7],
- application of transfer learning and fine-tuning of neural models to accelerate research.

Acknowledgment

The research presented in this paper was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project VEGA 1/0753/20, KEGA 009TUKE-4/2019 and by Slovak Research and Development Agency project APVV-15-0517.

References

[1] J. R. Pierce, "Whither speech recognition?" *The Journal of the Acoustical Society of America*, vol. 46, no. 4B, pp. 1049–1051, Oct. 1969. [Online]. Available: https://doi.org/10.1121/1.1911801
[2] P. Viszlay, J. Staš, T. Koctúr, M. Lojka, and J. Juhár, "An extension of the slovak broadcast news corpus based on semi-automatic annotation," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France: ELRA, May 2016.
[3] H. Chen, S. Lundberg, and S.-I. Lee, "Checkpoint ensembles: Ensemble methods from a single training process," 2017.
[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.
[5] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *CoRR*, vol. abs/1409.0473, 2014.
[6] S. Toshniwal, A. Kannan, C.-C. Chiu, Y. Wu, T. Sainath, and K. Livescu, "A comparison of techniques for language model integration in encoder-decoder speech recognition," in *IEEE Spoken Language Technology Workshop 2018*, 12 2018, pp. 369–375.
[7] E. McDermott, H. Sak, and E. Variani, "A density ratio approach to language model fusion in end-to-end automatic speech recognition," 2020.

Cryptography accelerator using symmetric and asymmetric algorithms

[1]*Michal HULIČ (4th year)*
Supervisor: [2]Liberios VOKOROKOS

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

[1]michal.hulic@tuke.sk, [2]liberios.vokorokos@tuke.sk

***Abstract*—This article deals with the issue of cryptography with accelerators using symmetric and asymmetric algorithms. The work combines two seemingly different industries. One of these industries is cryptography, a scientific discipline dealing with the secrecy of confidential information from unwanted persons. The second industry is hardware accelerators, devices that are designed to perform some functions more efficiently than software. Based on the information in chapter three, measurements of the efficiency of the implementation of the cryptographic algorithm were performed using a hardware accelerator.**

***Keywords*—Cryptography, encryption, decryption, computation, symmetric cryptography, asymmetric cryptography.**

I. Introduction

A way how to secure modern communication technologies so that they can guarantee that the information that is transmitted reaches only the people for whom it is actually intended and no one else, so the issue of information security is solved by science, cryptography [1].

Cryptography deals with the implementation of security cryptographic mechanisms, whether in software or hardware. Software implementation of security cryptographic mechanisms is the implementation of mechanisms at the kernel level of the operating system, in operating system modules that use multiple applications, or directly at the application level, using program modules. The hardware implementation of security mechanisms deals with cryptographic tokens, which include cryptographic smart cards, USB tokens and cryptographic modules – accelerators.

Hardware security devices, sometimes referred as cryptographic coprocessors or cryptographic modules, are security devices designed to securely perform cryptographic operations in an otherwise untrusted environment. These may include smart cards used to authenticate or store sensitive data or cryptographic accelerators used to speed up cryptographic operations [2].

Due to microprocessor operations are often complex and microprocessor resources are limited, these devices require a coprocessor in addition to the microprocessor. A cryptographic accelerator is a coprocessor designed specifically to perform demanding mathematical and cryptographic operations to handle them better than standard CPUs without specialized focus [3].

II. DES and ElGamal algorithms

A. Symmetric encryption and its characteristic

Considering an encryption scheme consisting of a set of encryption transformations $\{E_e \in K\}$ and decryption transformations $\{D_d \in K\}$, where K is a set of keys. A symmetric key is a key {e, d} for which it holds that for each e it is easy to calculate its corresponding d only with knowledge e and for each d it is easy to calculate its corresponding e only with knowledge d. Since in most cases e = d holds e = d, we call these keys symmetric. Symmetric key encryption is called symmetric encryption [4].

Let $\{E_e \in K\}$ be a set of cryptographic transformations and $\{D_d \in K\}$ be a set of decryption transformations, where K is a set of keys. Consider any pair consisting of an encryption and decryption transformation (E_e, D_d) and assume that for each pair it is impossible to calculate D_d based on E_e alone. Let us have a random cipher c C, with which we will want to find the message m M such that E_e (m) = c. This property implies that it is impossible to calculate the corresponding d from a given c [4].

For this article the DES algorithm will be used to show the comparison between symmetric and asymmetric ways of computation. The DES algorithm is presented at [5].

B. Asymmetric encryption and its characteristic

As with symmetric encryption, consider an encryption scheme consisting of a set of encryption and decryption transformations $\{E_e \in K\}$ and $\{D_d \in K\}$, where K is a set of keys. Assume that each pair is associated with an encryption / decryption transformation (E_e, D_d) and assume that each pair has an E_e knowledge that is impossible to compute. Let us have a random cipher c C, with which we want to find the message m M, such that E_e (m) = c. This property implies that it is impossible to calculate the corresponding d from a given e [6][9].

ElGamal is an algorithm used in cryptography for asymmetric encryption [7]. ElGamal Public Key Encryption Algorithm: A public key and a corresponding private key are created for each participant. The following operations must be performed by the participant communication of the Alliance:

1. generation of a random large prime number p and a generator for a multiplicative group Z_p^*, from the set of integers modulo p.
2. selection of a random integer $a, 1 \leq a \leq p-2$ and calculation and mod p.

3. public key is a trinity (p, α, α^a), where a is considering as private key.

III. Comparison of Algorithms

In this chapter the measured values above the DES and ElGamal algorithms are presented. The values given for a given length are the average of 20 measurements. At the input of these 20 measurements, the length of the plaintext was maintained, but the characters were varied to try different combinations of English alphabet characters, numbers and special characters to take into account the variety of inputs from the alphabet of the word.

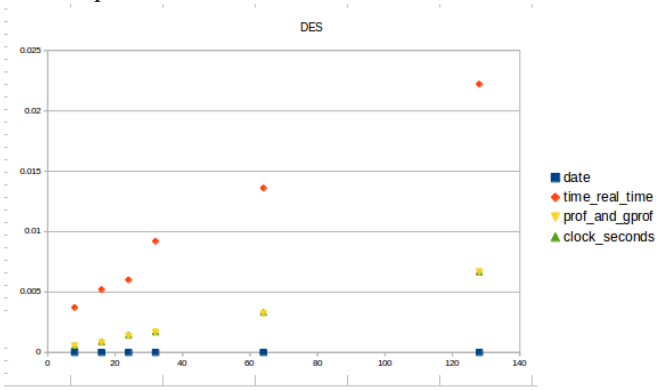


Fig. 1. The results of computation using DES algorithm

The same applies to the term Plain length as in the previous chapter. In order to be able to compare the algorithms, the same measurement methods, the same number of measurements and the same program inputs as in the previous subchapter had to be used. Thus, even with the following measured values, they are an average of 20 measurements. At the time of entry into these 20 measurements, the plaintext length was maintained, and the characters were varied to test different combinations of English alphabet characters, numbers, and special characters.

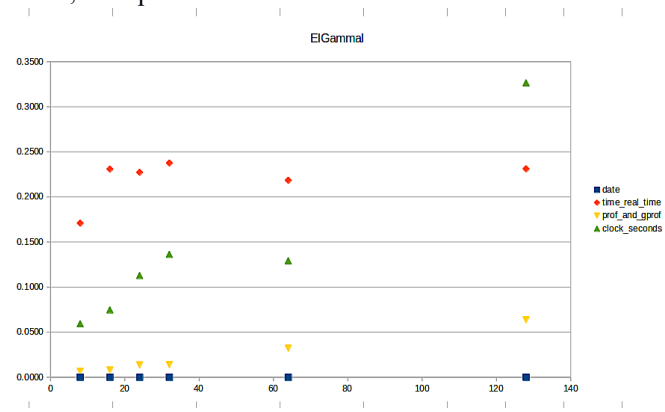


Fig. 2. The results of computation using ElGamal algorithm.

Presented graphs represents the dependence of the measured time on the plaintext length. For measurement different methods were used [8]:

1. date command
2. time command (the table shows the real time value)
3. prof and gprof
4. clock ().

As with the DES algorithm, the ElGamal algorithm has confirmed that as the length of the encrypted message increases, so does the encryption execution time.

Consider the prof and gprof method to be the authoritative method, as with the DES algorithm. Measurements showed that encrypting a 128-character message took 0.00666s. If we take into account these values, we get the function of the dependence of the encryption time on the input length l: t (l) = (496 · 106) · l + c, where the parameter l denotes the number of plaintext characters and c denotes the constant that is added to the function due to possible measurement deviation. This is a linear function.

IV. Conclusion

The result of this publication was to prove that the algorithm of symmetric cryptography DES is approximately 10 times faster than the algorithm of El Gamal, which is in line with the knowledge gained in the study of these algorithms. Among a number of cryptographic algorithms, the symmetric DES encryption algorithm and the asymmetric encryption algorithm were chosen El Gamal algorithm.

Thanks to the performance of today's computers, even with the longest tested inputs, it was only a difference of the order of hundreds of seconds.

The second result of the work was to verify the functionality and efficiency of the implementation of the DES algorithm using a hardware accelerator. For verification purposes, an implementation in VHDL language using FPGA technology was chosen. In this task, we found that for any length of plaintext, the measured time over the implementation of the DES algorithm in C ++ is higher than in VHDL, that confirms the results presented in the figures. VHDL technology has significant impact to the consequence in increase in computation performance.

Acknowledgment

This work was supported by the Faculty of Electrical Engineering and Informatics at the Technical University of Košice under contract No. FEI-2021-76: A modern interpreter of predicate linear logic formulas.

References

[1]J. Krhovjak, D. Cvrcek, V. Matyas. “Hardverové bezpecnostní moduly”. 2004. Fakulta informatiky Masarykovy univerzity v Brně.
[2]W. Stallings, L. Brown. 2012. “Computer Security Principles and Practice”, Second Edition. Prentice Hall PressOne Lake Street Upper Saddle River, NJUnited States. 978-0-13-377392-7.
[3]A. Menezes, P. C. Oorchot, S. A. Vanstone. “Handbook of Applied Cryptography”. 1996. CRC Press ISBN: 0-8493-8523-7
[4]A. Agath, C. Sidapara, D. Upadhyad. “Critical Analysis of Cryptography and Steganography”. 2018. International journal of scientific research in science, engineering and technology. ISSN 2394-4099 |
[5]G. Rouvroy, F. X. Standaert, J. J. Quisquater, J. D. Legat, d’ Efficient Uses of FPGAs for Implementations of DES and Its Experimental Linear Cryptanalysis, IEEE TRANSACTIONS ON COMPUTERS, VOL. 52, NO. 4, APRIL 2003.
[6]S. Khaled, H. Hussien, S. Yehia, d’FPGA Implementation of RSA Encryption Algorithm for E-Passport Application, World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:8, No:1, 2014
[7]P. Vishwanath, R. C. Joshi, A. K. Saxena, Fpga Implementation of Des Using Pipelining Concept With Skew Core Key-Schedulingē. Journal of Theoretical and Applied Information Technology, 2009.
[8]A. C. Shantilal, A Faster Hardware Implementation of RSA Algorithm, Department of Electrical & Computer Engineering, Oregon State University, Corvallis, Oregon 97331 USA.
[9]A. H. Ansari, A. R. Landge, RSA algorithm realization on FPGA, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 7, July 2013.

Position-Sensorless Control of Permanent Magnet Synchronous Motor

[1]*Viktor Petro (1st year),*
Supervisor: [2]*Karol Kyslan*

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

[1]viktor.petro@tuke.sk, [2]karol.kyslan@tuke.sk

***Abstract*—This paper reveals an overview of sensorless control of the permanent magnet synchronous motor. At first, approaches based on the motor model will be classified and the pros and cons of the methods will be discussed in detail. An explanation of the sliding mode observer for back electromotive force observation will be proposed. Furthermore, multiple HF signal injection based methods will be classified. Different operation regions and applications for the individual approaches are presented. The current state of the research and problems are briefly discussed and the future direction of the work is given in the conclusion.**

***Keywords*—Permanent magnet synchronous motor, sensorless control, sliding mode observer, HF signal injection**

I. Introduction

Permanent magnet synchronous motors (PMSM) are nowadays widely used in industry and home applications due to their high power density, dynamic performance, wide speed range, efficiency, and reliability. Field oriented control (FOC) with position sensor feedback in velocity control loop has become a standard control strategy for most of the applications. The basic idea behind the FOC is decoupling of the torque and flux producing component of the stator winding currents. This decoupling allows to control the complex three-phase PMSM in the same principle as a classical DC motor with separate excitation. In the case of a separately excited DC machine, the armature winding current is used for torque generation and the excitation current for the flux generation. Since the SMPM has permanent excitation provided by the permanent magnets, the excitation current is usually set to zero. Although under the flux weakening operation it can be set to negative values. For the FOC a transformation into the rotor reference frame is used for obtaining the direct and quadrature axes decoupled values. This transformation is called the Park transformation for which the correct electrical rotor position information is essential. Usually, position sensors directly coupled to the rotor shaft provide precise rotor position information in full speed range. Thus a smooth speed control can be achieved from zero to nominal speed. However, using a position sensor decreases the overall reliability and robustness of the drive system. Among that, eliminating the position sensor brings further more advantages, such as cost reduction, drive compactness, increased mechanical robustness, etc. On the other hand, reduced preciseness and extra cost for the software development needs to be concerned in sensorless drive applications.

II. Sensorless control of PMSM

In sensorless control, rotor shaft position and speed are not measured but instead of being measured, they are estimated with the use of electrical variables such as stator windings currents or voltages. In the last few decade enormous efforts have been made in the research of sensorless PMSM control. The conventional estimation methods can be classified into three main groups [1]:

- fundamental excitation model-based methods,
- saliency and signal injection-based methods,
- artificial intelligence-based methods.

The fundamental excitation model-based methods are mainly using the electromotive force (EMF) and this category includes the Luenberger observer, the Kalman filter, the flux observer, the sliding-mode observer (SMO), estimation based on the model reference adaptive system (MRAS), etc. The SMO is widely used in industrial applications due to its robustness against motor parameter variations and disturbances. For surface mounted permanent magnet synchronous machine (SMPMSM) a classical second-order SMO in the stationary reference frame (SRF) $\alpha\beta$ can be used [2] – [6] for indirect back-EMF estimation. In [7] a modified second order SMO was introduced based on an extended EMF model which can be also used for interior permanent magnet synchronous motor (IPMSM). Another option for sensorless control of IPMSM was proposed in [8], where a second-order SMO was designed in rotor reference frame (RRF) dq. In [9] – [12] a modified high-order SMO was proposed for direct back-EMF estimation. In this case, the back-EMF components stand as state variables. To suppress the well-known chattering problem of the SMO based on signum function, saturation [4] or sigmoid [13] functions were adopted instead.

The mentioned model-based methods have the advantage of simplicity and straightforwardness. However, they provide reliable rotor position and speed information only above 3% of nominal speed [14], since the back-EMF amplitude is low or even zero at low-speed range and standstill. Thus model based methods can be applied only for medium and high speed range sensorless control. For low speed and standstill sensorless drive control, signal injection-based methods and fundamental PWM excitation based methods were developed. High frequency (HF) signal injection is usually preferred instead of low frequency (LF) signal injection. The injected signal can be injected either in stationary reference frame or in the observed rotor reference frame. Furthermore, the

injected signal can be superimposed onto the fundamental frequency signal required for the motor control or can be injected independently into the stator windings while the FOC is interrupted. Commonly used HF injection-based methods can be further classified:

- HF sinusoidal signal injection-based method,
  - HF rotating sinusoidal signal injection,
  - HF pulsating sinusoidal signal injection,
- HF square-wave signal injection-based method,
- HF pulse signal injection-based method.

Hybrid sensorless methods were proposed for operation in full speed range [15]–[17]. In those, both model-based and signal injection-based methods are combined.

In the case of soft computing based technique (or artificial intelligence technique) the accuracy and reliability of the observation of rotor position is independent on the speed range. They are more likely dependent of the relevant choice of input data which affects the observation performance [18].

A. Sling mode observer

The SMO is wildly used in industrial applications due to its robustness. Texas Instruments offers for customers software library with SMO for digital microcontrollers described in [2]. Microchip has also developed a sensorless control based on the SMO for electrical drives [3]. Whirlpool has developed a sliding mode sensorless control of PMSM for washing machine applications [4]. The mentioned companies implemented a second-order SMO. In this case, the observer treats the back-EMF components as disturbances. NXP is using SMO in compressor control and other consumer and industrial applications [8]. In this case, a full-order observer is used where the back-EMF voltages are considered as state variables. Similarly in [10] a full order SMO is proposed with an additional adaptive speed scheme.

In this article a 2^{nd} order SMO based on extended EMF E_{ex} similar to one in [9] will be proposed. The mathematical model for PMSM sliding mode position observer is most of the time constructed in the stationary $\alpha\beta$ reference frame. The state equation where the stator current components are state variables are:

$$\begin{bmatrix}\dot{i}_\alpha\\ \dot{i}_\beta\end{bmatrix} = \frac{1}{L_d}\begin{bmatrix}-R_s & -\omega_e(L_d-L_q)\\ \omega_e(L_d-L_q) & -R_s\end{bmatrix}\begin{bmatrix}i_\alpha\\ i_\beta\end{bmatrix} + \frac{1}{L_d}\begin{bmatrix}u_\alpha-e_\alpha\\ u_\beta-e_\beta\end{bmatrix}, \tag{1}$$

where u_α, u_β, i_α, and i_β are the stator voltage and current components in SRF respectively, L_d and L_q are the direct and quadrature axis inductance respectively, ω_e is the electrical angular velocity, $R_s = R_{2ph}/2$ where R_{2ph} is the resistance between two phases and $e_\alpha = -E_{ex}\sin(\theta_e)$ and $e_\beta = E_{ex}\cos(\theta_e)$ are the back-EMF components, where θ_e is the electrical rotor position and $E_{ex} = \omega_e\tau_{PM}$ is the extended EMF, where:

$$\tau_{PM} = \frac{2}{3}\frac{k_t}{p} = \frac{k_e}{p}, \tag{2}$$

is the permanent magnet flux, where k_t is the motor torque constant, k_e is the EMF constant and p is the motor pole pairs number. In the case of SMPMSM where we assume that $L_d = L_q$ equation (1) is simplified and for the inductances stand: $L_d = L_q = L_s = L_{2ph}/2$, where L_{2ph} is the inductance between two phases. Since during the position-sensorless motor control only the stator currents are available, the switching function for the SMO is defined as follows:

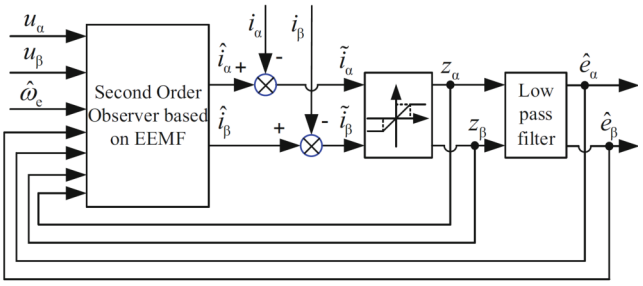


Fig. 1. Block diagram of the 2^{nd} order SMO [9] .

$$\mathbf{s} = \begin{bmatrix}\bar{i}_\alpha\\ \bar{i}_\beta\end{bmatrix} = \begin{bmatrix}\hat{i}_\alpha - i_\alpha\\ \hat{i}_\beta - i_\beta\end{bmatrix}, \tag{3}$$

where superscript '^' indicates that the parameter is observed. The sliding surface also referred to as a hyperplane is selected as:

$$\mathbf{s} = 0. \tag{4}$$

The matrix form of a conventional SMO is shown below:

$$\begin{bmatrix}\dot{\hat{i}}_\alpha\\ \dot{\hat{i}}_\beta\end{bmatrix} = \frac{1}{L_d}\begin{bmatrix}-R_s & -\hat{\omega}_e(L_d-L_q)\\ \hat{\omega}_e(L_d-L_q) & -R_s\end{bmatrix}\begin{bmatrix}\hat{i}_\alpha\\ \hat{i}_\beta\end{bmatrix} + \frac{1}{L_d}\begin{bmatrix}u_\alpha-\hat{e}_\alpha-z_\alpha\\ u_\beta-\hat{e}_\beta-z_\beta\end{bmatrix}, \tag{5}$$

where z_α and z_β are SMO feedback components:

$$\begin{bmatrix}z_\alpha\\ z_\beta\end{bmatrix} = k\begin{bmatrix}\text{sign}(\hat{i}_\alpha - i_\alpha)\\ \text{sign}(\hat{i}_\beta - i_\beta)\end{bmatrix}, \tag{6}$$

where parameter k is the SMO gain and according to the Lyapunov stability analysis [19]:

$$k > \max\left(|e_\alpha|, |e_\beta|\right). \tag{7}$$

The estimated components of the back-EMF are obtained using a low-pass filter (LPF):

$$\begin{bmatrix}\hat{e}_\alpha\\ \hat{e}_\beta\end{bmatrix} = \frac{\omega_c}{s+\omega_c}\begin{bmatrix}z_\alpha\\ z_\beta\end{bmatrix}, \tag{8}$$

where ω_c represents the cut-off frequency of the LPF. The block diagram of the SMO is shown in Fig. 1. Finally the electrical rotor position can be obtained as flows:

$$\hat{\theta}_e = -\arctan\left(\frac{\hat{e}_\alpha}{\hat{e}_\beta}\right). \tag{9}$$

For electrical angular speed stands:

$$\hat{\omega}_e = \frac{d\hat{\theta}_e}{dt}. \tag{10}$$

Since the LPF was used to obtain the estimated back-EMF components in eq. (8) a phase lag is introduced in the observed electrical rotor position (9). Therefore $\hat{\theta}_e$ needs to be compensated as follows:

$$\hat{\theta}_{eCOMP} = \hat{\theta}_e + \Delta\hat{\theta}_e = \hat{\theta}_e + \arctan\left(\frac{\hat{\omega}_e}{\omega_c}\right). \tag{11}$$

The observed rotor position after compensation (11) can be finally used in field-oriented control as the feedback value instead of its measurement by a sensor.

B. HF signal injection based methods

The rotating (also called revolving) HF sinusoidal signal injection based method was first proposed by R. D. Lorenz [20]. This technique is also often called as "$\alpha\beta$" injection since the rotating voltage signal is injected in the SRF. For two orthogonal voltage signals with constant amplitude rotating at the injected frequency stands:

$$\begin{bmatrix} u_{\alpha h} \\ u_{\beta h} \end{bmatrix} = U_h \begin{bmatrix} \cos(\omega_h t) \\ \sin(\omega_h t) \end{bmatrix}, \tag{12}$$

where subscript h represents HF components, U_h is the amplitude, and ω_h is the frequency of the injected HF voltage. The block diagram of the HF rotating sinusoidal signal injection based method is shown in Fig. 2. The superimposed HF

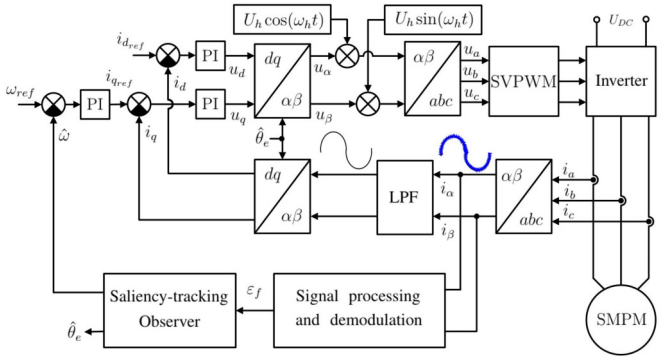


Fig. 2. Block diagram of rotating HF sinusoidal signal injection based method.

voltage signal onto the fundamental frequency results in HF induced current. According to [21], only the negative sequence of the induced current component contains rotor position information. To extract this information, a heterodyning process and a saliency-tracking observer can be used. First the HF current component needs to be separated from fundamental frequency phase current. For those purposes, several techniques were proposed [7], [22]. In [9] a high-pass filter (HPF) is used for HF current extraction. Then during the heterodyning process, the HF currents are multiplied by $\cos(2\hat{\theta}_e - \omega_h t)$ and $\sin(2\hat{\theta}_e - \omega_h t)$. Then the sum of the two HF currents in SRF is filtered using a low-pass filter and marked as ε_f. This signal is proportional to position estimation error, i.e. $\varepsilon_f \approx \tilde{\theta}_e$, where $\tilde{\theta}_e = \theta_e - \hat{\theta}_e$. Furthermore, ε_f is used as the input for the saliency-tracking observer which is a Luenberger style position and velocity observer [20], [23]. The task of the observer is to drive ε_f to zero and suppress the estimation error $\tilde{\theta}_e$ resulting into the convergence of the estimated position to the real one, i.e. $\hat{\theta}_e \rightarrow \theta_e$. The main disadvantage of the method is current fluctuation and torque ripples caused by the induced HF current. Furthermore, HPF and LPF are necessary for signal extractions whereas the LPF restricts the bandwidth of controllers.

To suppress the torque ripples a different approach was developed which relies on HF pulsating sinusoidal signal injection. In this case, the signal is injected in the observed RRF and for the injected voltage stands [24]:

$$\begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} = \begin{bmatrix} U_h \sin(\omega_h t) \\ 0 \end{bmatrix}. \tag{13}$$

As can be seen in eq. (13), the HF pulsating voltage is only injected into the observed d axis of RRF. Furthermore, only $i_{\hat{q}h}$ is required in the heterodyning process where it is multiplied by $\cos(\omega_h t)$. After filtering this product with a LPF, signal ε_f is obtained which is associated with the position estimation error $\tilde{\theta}_e$. Further, ε_f is used as an input for the saliency-tracking observer. Another approach is presented in [9], where the induced HF currents are measured in a new reference frame $d^m q^m$ which lags behind the estimated RRF by $\pi/4$. In both mentioned cases the HF voltage is injected into the d axis of RRF decreasing the moment ripples in comparison with the HF rotating sinusoidal signal injection-based method. However, the mutual disadvantage of both methods is frequency limitation, since the injected HF signal needs to remain sinusoidal. Therefore, the dynamic performance of the observation is limited.

To overcome these limitations the HF square-wave signal injection-based approach was developed where the frequency of the injected signal can be higher and therefore offers better dynamic performance [25]. Similarly as in the case of HF pulsating sinusoidal signal injection-based method, the HF square-wave signal is injected into the observed RRF and is defined as follows:

$$\begin{bmatrix} u_{\hat{d}h} \\ u_{\hat{q}h} \end{bmatrix} = \begin{bmatrix} U_h(-1)^k \\ 0 \end{bmatrix}, \tag{14}$$

where U_h is the amplitude and k is the sequence of the injected voltage. Since a higher frequency of the injected signal can be achieved, the LPF in the current feedback loop can be omitted. It results in higher bandwidth of the current controllers compared to HF sinusoidal injection-based methods. To solve the inaccuracy in observed position caused by the voltage errors over the inverter and stator windings, negative $-U_h$ and positive U_h voltage components were injected in approach proposed in [26]. This resulted in the necessity of two current sampling periods and thus delaying the control process. These methods bring overall improvement of the sensorless control dynamic performance although high power losses needs to be concerned.

All the aforementioned HF signal injection-based approaches work parallel with FOC and thus the injected signal is superimposed onto the fundamental frequency component. Therefore the induced HF current needs to be extracted from the measured current using a HPF, or in some applications a band-bass filter (BPF). Further, the LPF may be also required in signal processing for the rotor speed and position estimation decreasing the bandwidth and dynamic performance of the drive. To eliminate the mentioned drawbacks a new approach based on HF pulse signal injection was proposed in recent years [27]. The idea behind this approach is to separate FOC from the injection. It means that the FOC is interrupted, while the injection is performed. By this approach, no digital filters are required for HF induced current extraction from the fundamental frequency component. When the injection is enabled a positive voltage component U_h is injected into the $\hat{d}$ axis while the injected voltage into the $\hat{q}$ axis is kept on zero. Similarly as in the previous method, to overcome the voltage errors, two voltage signals U_h and $-U_h$ may be injected. Then a normalization process is followed what results in the signal ε which is associated with position estimation error. It acts as an input for the saliency-tracking observer. For all the mentioned approaches the same observer can be used described in [23] or a PLL tracking scheme used in [12].

III. Conclusion

A state-of-art for mid-to-high speed estimation methods and low-speed estimation methods were presented. The advantages of such control are obvious but some of the disadvantages need to be mentioned too. One of them is the lack of full-speed range operation. The methods based on the back-EMF observation fail to estimate the position reliably below a certain speed. Usually, an open-loop start-up procedure is provided until the amplitude of the back-EMF became observable. During this procedure the stator and rotor fluxes are not aligned well, therefore the motor cannot be loaded unpredictably otherwise the PMSM may lose synchronization and stops its movement. To solve the problem of unobservability of the PMSM at low and zero speed different injection based methods were proposed. The main issue is current fluctuation and torque ripples introduced by the HF injected signal. To overcome this limitation an approach with FOC interrupt was proposed with voltage errors compensation. But still, the problem of full-speed range control is not guaranteed using only one of the mentioned methods. Therefore hybrid schemes are often used for combining the model-based and injection-based methods. And so the future research work is concerned with a combination of the approaches to find the optional solution for full range speed and torque control of the PMSM in sensorless application. It should be stressed that all sensorless methods for the control of PMSM should be developed as a purely software-based algorithm. From this point of view, we will not assume any special type of sensors such as acceleration sensors or current-derivative sensors. This premise is completely in line with industrial requirements.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic under the project VEGA 1/0493/19.

References

[1] 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.

[2] T. INSTRUMENTS, "Digital Motor Control Software Library.pdf," no. October, 2003. [Online]. Available: https://www.ti.com/lit/ug/spru485a/spru485a.pdf

[3] Microchip, "AN1078-Sensorless Field Oriented Control of a PMSM," *an1078*, vol. PP, no. 1, pp. 1–6, 2019. [Online]. Available: http://ww1.microchip.com/downloads/en/appnotes/01292a.pdf

[4] S. Chi, Z. Zhang, and L. Xu, "Sliding-mode sensorless control of direct-drive PM synchronous motors for washing machine applications," *IEEE Transactions on Industry Applications*, vol. 45, no. 2, pp. 582–590, 2009.

[5] V. Utkin, J. Guldner, and J. Shi, *Sliding mode control in electro-mechanical systems, second edition*, 2017. [Online]. Available: https://doi.org/10.1201/9781420065619

[6] C. Li and M. Elbuluk, "A sliding mode observer for sensorless control of permanent magnet synchronous motors," *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*, vol. 2, no. C, pp. 1273–1278, 2001.

[7] Z. Chen, M. Tomita, S. Doki, and S. Okuma, "An extended electromotive force model for sensorless control of interior permanent-magnet synchronous motors," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 2, pp. 288–295, 2003.

[8] Freescale Semiconductor, "Sensorless PMSM Vector Control with a Sliding Mode Observer for Compressors Using MC56F8013," 2008. [Online]. Available: https://www.nxp.com/docs/en/reference-manual/DRM099.pdf

[9] G. Wang, G. Zhang, and D. Xu, *Position sensorless control techniques for permanent magnet synchronous machine drives*, 2019. [Online]. Available: https://doi.org/10.1007/978-981-15-0050-3

[10] Z. Chen, M. Tomita, S. Doki, and S. Okuma, "New adaptive sliding observers for position-and velocity-sensorless controls of brushless dc motors," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 582–591, 2000.

[11] F. Parasiliti, R. Petrella, and M. Tursini, "Sensorless speed control of a PM synchronous motor based on sliding mode observer and extended Kalman filter," *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*, vol. 1, pp. 533–540, 2001.

[12] G. Wang, Z. Li, G. Zhang, Y. Yu, and D. Xu, "Quadrature PLL-based high-order sliding-mode observer for IPMSM sensorless control with online MTPA control strategy," *IEEE Transactions on Energy Conversion*, vol. 28, no. 1, pp. 214–224, 2013.

[13] H. Liu and G. Li, "Rotor Position and Speed Estimation Method for Magnetically Suspended PMSM Based on Modified Sliding Mode Observer," no. Mcae, pp. 210–214, 2016.

[14] Z. Ma and R. Kennel, "System-on-Chip sensorless control of PMSM combining signal injection and flux observer," *Conference Proceedings - 2012 IEEE 7th International Power Electronics and Motion Control Conference - ECCE Asia, IPEMC 2012*, vol. 2, pp. 1201–1205, 2012.

[15] C. Silva, G. M. Asher, and M. Sumner, "Hybrid rotor position observer for wide speed-range sensorless PM motor drives including zero speed," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 2, pp. 373–378, 2006.

[16] A. Piippo, M. Hinkkanen, and J. Luomi, "Analysis of an adaptive observer for sensorless control of interior permanent magnet synchronous motors," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 2, pp. 570–576, 2008.

[17] H. Iura, M. Inazumi, T. Kamei, and K. Ide, "Hybrid sensorless control of IPMSM for direct drive applications," *2010 International Power Electronics Conference - ECCE Asia -, IPEC 2010*, pp. 2761–2767, 2010.

[18] M. Larrea, E. Irigoyen, V. Ǵomez, and F. Artaza, "Nonlinear system control based on neural networks with adaptive predictive strategy," in *2010 IEEE 15th Conference on Emerging Technologies Factory Automation (ETFA 2010)*, 2010, pp. 1–7.

[19] M. Elbuluk and C. Li, "Sliding Mode Observer for Wide-Speed Sensorless Control of PMSM Drives," *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*, vol. 1, pp. 480–485, 2003.

[20] P. L. Jansen and R. D. 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.

[21] D. D. Reigosa, P. García, D. Raca, F. Briz, and R. D. Lorenz, "Measurement and adaptive decoupling of cross-saturation effects and secondary saliencies in sensorless controlled IPM synchronous machines," *IEEE Transactions on Industry Applications*, vol. 44, no. 6, pp. 1758–1767, 2008.

[22] M. Koteich, A. Messali, and S. Daurelle, "Self-sensing control of the externally-excited synchronous machine for electric vehicle traction application," *Conference Proceedings - 2017 8th International Symposium on Sensorless Control for Electrical Drives, SLED 2017*, no. 6, pp. 91–96, 2017.

[23] K. W. Van, "High-Resolution Velocity Estimation for All-Digital, ac Servo Drives," *IEEE Transactions on Industry Applications*, vol. 27, no. 4, pp. 701–705, 1991.

[24] D. Raca, P. García, D. D. Reigosa, F. Briz, and R. D. Lorenz, "Carrier-signal selection for sensorless control of PM synchronous machines at zero and very low speeds," *IEEE Transactions on Industry Applications*, vol. 46, no. 1, pp. 167–178, 2010.

[25] D. Kim, Y. C. Kwon, S. K. Sul, J. H. Kim, and R. S. Yu, "Suppression of injection voltage disturbance for high-frequency square-wave injection sensorless drive with regulation of induced high-frequency current ripple," *IEEE Transactions on Industry Applications*, vol. 52, no. 1, pp. 302–312, 2016.

[26] 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.

[27] 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.

Speech activity detection with visual stimuli

[1]*Marianna KOCTÚROVÁ (4th year),*
Supervisor: [2]*Jozef JUHÁR*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]marianna.kocturova@tuke.sk, [2]jozef.juhar@tuke.sk

***Abstract*—In this paper research on speech detection using electroencephalographic signals is described, which carries information about speech brain activity and brain activity induced by visual stimuli. The research was performed with four experimental subjects, which were exposed to visual stimulus to stimulate the speech. The research uses the Feed-forward Neural Network classification algorithm, which was able to create a speech detection model using only Electroencephalographic signals with an F1 score of 73.30 %. We believe that our scientific contribution can lead to a better BCI-based speech recognizer.**

***Keywords*—Electroencephalography, Speech detection, Feed-Forward Neural Network, Visual stimuli**

I. Introduction

This paper presents the final part of the research of the dissertation thesis and its outputs. The presented work deals with the speech detection from Electroencephalographic (EEG) signals, which does not require audio input data. The examined EEG data included brain electrical activity from speech processes and brain activity induced by visual stimulation. This type of EEG data should contain information about such brain activity that is induced in common speech communication. We consider that human communication is based on the creation of words but also on the creation of visual images in the human brain. The relationship between imagination and vision is very close and these two actions elicit similar brain electrical activities [1], [2]. Image recognition and image naming involve the activity of occipitotemporal and prefrontal brain areas [3].

The used EEG data obtained from a mobile device, which better responds to the possibilities of using Brain-computer interface (BCI) technology in real conditions. Speech detection from EEG signals, reveals the speech brain signal, using classification algorithms. For this research Feed-forward neural network was used.

II. An Experimental Database

For reasons of this research were recorded EEG signals from 4 subjects. Each subject was instructed to follow an experimental protocol. EEG data was captured with mobile EEG device Ultracortex Mark III OpenBci [4]. The EEG electrode configuration on our device was set to record the brain activity in different areas of the head. The channels used were namely Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2, according to the international 10/20 system [5]. Figure 1a) shows the top view of the EEG electrodes configuration, 1b) shows the side view of the location of the electrodes used in our research and the iodine of the brain is shown. Especially we tried to cover the following areas:

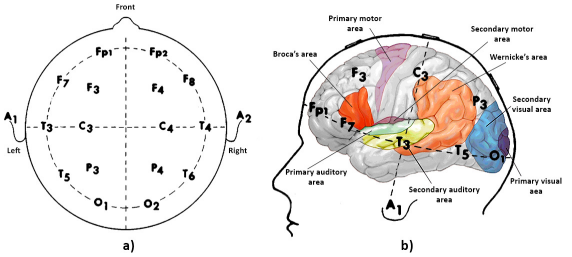


Fig. 1. a) Electrode configuration according to 10/20 international system used in the experiment. b) Lateral view of the electrode configuration, compared with the brain areas (Motor, Auditory, Visual, and Speech)

Speech areas (Broca's and Wernicke's area), Auditory areas, Visual areas and Motor areas.

A. Experimental Protocol

An experimental protocol for EEG recording was designed to activate the speech, visual and imaginary areas of the brain. EEG signals recording was performed in two sessions, each lasting about 10 minutes. In the first session specific colours were displayed on the screen as pictures, while in the second session, the colours were displayed as text on the screen. The recorded subject was tasked to focus on the displayed colours and to name them. In the above exercise, 10 colours were displayed randomly one by one with short breaks represented by a plain black screen. The colours selected for this protocol were yellow, green, red, blue, orange, violet, white, pink, grey, and brown.

The display of one color lasted 4 seconds, followed by a pause of 1 second to help the subject release cognitive functions for a while. The length of pronunciation of one word to the subject lasted on average 0.75 s and the time of pause between pictures when the subject did not speak lasted on average 4.25. The whole experimental protocol, even with timing, as shown on the Figure 2.

III. Training Algorithm

In our experiment, the 2-layer Feed-forward neural network was used. Min-max scaling was performed by network preprocessing. The neural network consists of a single tanh activated hidden and binary output sigmoid activated output layer. We trained the network with Scaled conjugate gradient backpropagation, with binary cross-entropy as the loss function. Output pseudo probabilities were thresholded with a 0.5 decision boundary. The first layer of the network consisted of one hundred hidden neurons.

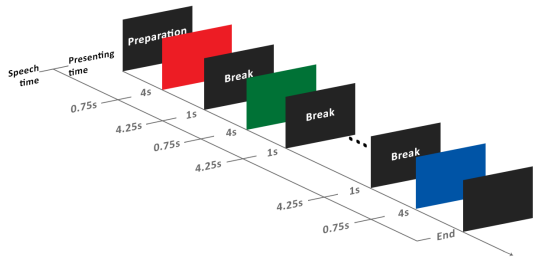


Fig. 2. Protocol for EEG signal acquisition. On-screen presentation times and average speech times. Individual colours or colour names were displayed with breaks created by black colour on screen.

In the experiment, a model was created using machine learning modelling, which would be able to recognize the state of speech production and the state of brain inactivity. Therefore, in the above classification, all words were explicitly considered as one class [6].

The experiment was divided into two main parts, in the first we tried to create single-subject models. In the second part, we created cross-subject models.

IV. Results

In this section, the results of neural network testing are presented. In the first part the speech detection models for individual subject data set were created. In the second part the focus was on the creation of a model from the cross-subject data set. One hundred models for each data set were trained using the assembled neural network. From all of the created models the one with the highest result in the testing process was selected and the results with the highest values for data set models are presented here.

The experimental results were evaluated using Accuracy and F1 score metrics. The aim of the research was to find positive results for the state of speech activity, indicated by output class 1, as the most important metric is considered to be the F1 score. The accuracy and the F1 score are defined by founded True Positive results (TP), True Negative results (TN), False Positive results (FP) and False Negative results (FN) by the formulas:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}. \quad (1)$$

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (2)$$

A. Single-subject Model Experiment

In the first part of the experiment, models for the single-subject models were created, 70% of the data was used for training, 15% for the model validation, and 15% for the test. As can be seen in Table I, machine learning learned to detect speech from the EEG signal for a single subject with an average accuracy of 84%. This accuracy can be considered relatively high.

B. Cross-subject Model Experiment

In this part we selected data sets from the combination of 3 subjects which were divided into 80% training set and 20% validation set. The created model was tested on the data set from the 4th subject.

TABLE I
Experimental results for test set Single-subject model

Subject number	Accuracy %	F1 score %
1	86.65	88.36
2	74.72	79.05
3	88.75	90.26
4	88.35	89.53
Average	84.62	86.80

TABLE II
Experimental results for test set Cross-session model

Subject numbers model training	Subject number model testing	Accuracy %	F1 score %
1, 2, 3	4	73.71	79.50
2, 3, 4	1	62.14	73.87
1, 3, 4	2	57.48	62.25
1, 2, 4	3	72.56	77.56
Average		66.47	73.30

Table II was compiled based on results for cross-subject speech detection. The data used for creating this detection model were mixed from all subjects and both sessions.

V. Conclusion

Table I shows the average result for single model accuracy of 84.6% and 86.8% of the F1 score. On the other hand, results for a more demanding model with cross data II shows that the best result was achieved for the configuration data set on subjects 1, 2 and, 3 in the training model and the testing the model from the data on subject 4. The F1 score here reached 79.5%. Although a small number of subjects was involved in the acquisition of the training data sets the relatively high accuracy of the results was achieved, which provides a basis for the development of research in this area.

Acknowledgment

This work was supported by Slovak Research and Development Agency project number APVV-15-0517, the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic project number VEGA 1/0753/20 and the Cultural and Educational Grant Agency of the Slovak Republic project number KEGA 009TUKE-4-2019.

References

[1] J. L. Breedlove, G. St-Yves, C. A. Olman, and T. Naselaris, "Generative feedback explains distinct brain activity codes for seen and mental images," *Current Biology*, 2020.

[2] C. I. Winlove, F. Milton, J. Ranson, J. Fulford, M. MacKisack, F. Macpherson, and A. Zeman, "The neural correlates of visual imagery: A co-ordinate-based meta-analysis," *Cortex*, vol. 105, pp. 4–25, 2018.

[3] M. Canini, P. A. Della Rosa, E. Catricalà, K. Strijkers, F. M. Branzi, A. Costa, and J. Abutalebi, "Semantic interference and its control: A functional neuroimaging and connectivity study," *Human brain mapping*, vol. 37, no. 11, pp. 4179–4196, 2016.

[4] O.-O. S. B. Tools, "Openbci.com. retrieved 24 february 2018," 2018.

[5] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[6] M. Koctúrová and J. Juhár, "A novel approach to eeg speech activity detection with visual stimuli and mobile bci," *Applied Sciences*, vol. 11, no. 2, p. 674, 2021.

Enhancing reinforcement learning agents with a priori knowledge using deictic representation

[1]*Lukáš HRUŠKA (3rd year),*
Supervisor: [2]*Peter SINČÁK*

[1,2]Dept. of cybernetics and artificial intelligence, FEI TU of Košice, Slovak Republic

[1]lukas.hruska@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—Most reinforcement learning algorithms operate directly on the propositional state representation. This results in poor generalization and strongly suffers from the curse of dimensionality. In this paper we look into a viable alternative - deictic representation. Moreover, we propose a method of incorporating external knowledge into the algorithm.**

***Keywords*—deep reinforcement learning, deictic representation, Gridworld, hierarchical reinforcement learning, Minecraft**

I. Introduction

Deep reinforcement learning ((D)RL) has emerged as a promising approach for autonomous acquisition of complex behaviours from low-level observation. Such algorithms mostly rely on a sandbox environment, where every mistake is of little consequence. However, this is not the case in the real world. Various approaches have been developed to speed up the learning by translating the human knowledge into the learning process. These include interactive RL [1][2] and learning from demonstration [3][4][5][6]. Other techniques aim at preprocessing the input, such as deictic [7][8][9] representation.

II. Deictic representation in reinforcement learning

The term originally taken from from Ancient Greek $\delta\epsilon\iota\kappa\nu\upsilon\mu\iota$ means to point out, to show. Hence, the idea behind deictic representation (DR) is to show relations between various objects instead of relying on the agent to learn relation on its own. It may be expressed in natural language as *the-PEN-I-am-pointing-at* or *the-DOOR-in-front-of-ME*. As such, we may mark important objects and discard the rest, rapidly simplifying the state space (one example of this is provided by [9] where the DR is able to reduce the number of possible states from 2^{30} down to 5184 - a 99.9995% reduction). This introduction of partial observability offers a huge trade-off [8] - it may speed up the process by obscuring irrelevant informations but it may also hinder it by obscuring relevant ones. Unlike propositional representation, which would need to distinguish each object, DR is able to generalize over similar objects - it does not matter whether I am holding *hammer2* or *hammer9* as long as they have similar qualities.

III. Experiments

In this section, we will look into the possibility of using deictic representation in conjunction with usual RL algorithms, such as DQN [10] or TRPO [11]. This is done by adding an additional layer in front of the algorithm. We used several GridWorld environments for various navigation tasks:

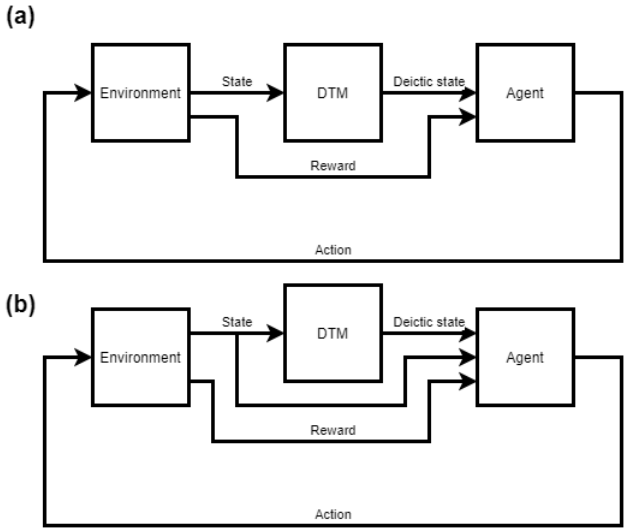


Fig. 1. Usual state matrix is being processed by Deictic Transformation Module (DMT). We investigate two ways of injecting the deictic representation into the reinforcement learning agent - pure deictic (a) where the agent receives only processed observation; and mixed deictic (b) where the agent is presented with a combination of both, the original and processed observation.

1) Navigation without obstacles, where the goal changes location after *n*-iterations.
2) Navigation without obstacles, where the goal changes location after each iteration.
3) Navigation with an enemy following the agent.

The $m \times n$ matrix representation of GridWorld is transformed using rule-based deictic transformation module (DTM) into a set of extracted objects, each described with 3 features - distance, angle and class (wall, goal, enemy) in relation to the agent. This replaces the original matrix (Fig.1a) or is added to it (Fig.1b). Network topologies used only differ in the input layer. Contrary to our expectation, pure deictic outperformed mixed input in these simple tasks and both outperformed the reference model. Further analysis in more complex environments is necessary to better evaluate the deictic representation.

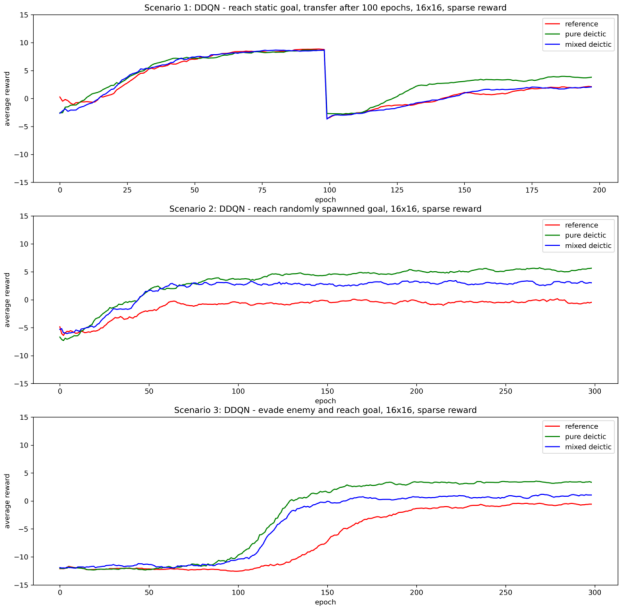


Fig. 2. Deictic proof of concept results: three types of agents (normal, pure deictic, mixed deictic) operate in a 16x16 Gridworld environment.

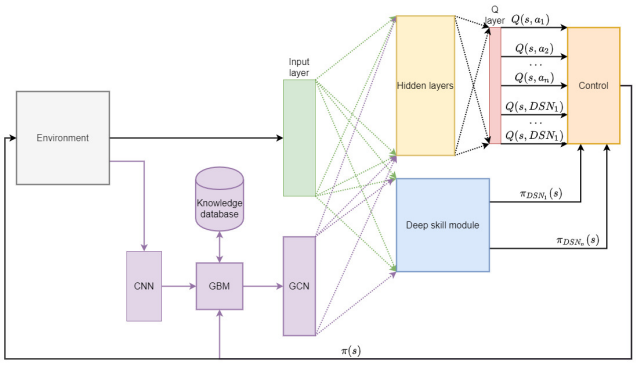


Fig. 3. Modified H-DRLN algorithm (figure based on [14], purple denotes new parts) - visual observation is obtained from a Minecraft-based environment which is then fed into convolutional network (CNN) for object recognition. Detected objects are then processed by the graph building module (GBM) to create a human-readable graph representation of the environment. This then serves as an additional input into the original network (or completely replaces the visual input), preprocessed with a graph convolutional network (GCN).

IV. Ongoing work

Minecraft is a well-known 3D video game with established machine learning community with platforms such as Malmo [12] and MineRL [13]. Due to its increasing levels of complexity, it is an ideal test-bed for hierarchical RL. Multiple algorithms have been developed in this environment, such as the one, we took as the basis: H-DRLN [14].

This algorithm uses two hierarchical layers of deep Q-networks - upper layer, composed of a single network, which is able to either select an action directly or pass the control down to one of the lower-level networks (called deep skill module) for finite number of steps. Each of these is pre-trained for a specific task. After the task is finished, the control is once again returned to the upper network.

Our proposed method (see Fig.3) aims to enhance this by encoding human knowledge into the input in form of a deictic representation. This should take the form of a graph with the central node representing the agent, other nodes objects, and edges relations among them. It is achieved by adding an additional convolutional network for object detection. Detected objects are then encoded into the aforementioned graph in the Graph builder module (GBM). GBM also keeps record of previous actions the agent undertook in an attempt to discover new relations. The goal is to analyse the impact of external knowledge on the quality of learning and performance of various layers in the hierarchical RL algorithm.

V. Conclusion

Deictic representation is often overlooked means to directly incorporate external knowledge into the input. We have provided an analysis of the topic and its application in the Gridworld environment. Moreover, we have analysed an existing algorithm and proposed a modification for deep deictic hierarchical reinforcement learning. Further research into this topic promises to yield usable solution.

Acknowledgment

This research is supported by AI4EU project from the European Union's Horizon 2020 research and innovation programme under grant agreement 825619 2019-2021.

References

[1] F. Cruz, S. Magg, C. Weber, and S. Wermter, "Training agents with interactive reinforcement learning and contextual affordances," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 4, pp. 271–284, 2016.

[2] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2285–2294.

[3] M. E. Taylor, H. B. Suay, and S. Chernova, "Integrating reinforcement learning with human demonstrations of varying ability," in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 617–624.

[4] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *arXiv preprint arXiv:1806.06877*, 2018.

[5] B. Piot, M. Geist, and O. Pietquin, "Boosted and reward-regularized classification for apprenticeship learning," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1249–1256.

[6] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 627–635.

[7] S. Finney, N. H. Gardiol, L. P. Kaelbling, and T. Oates, "Learning with deictic representation," 2002.

[8] S. Finney, N. Gardiol, L. P. Kaelbling, and T. Oates, "The thing that we tried didn't work very well: Deictic representation in reinforcement learning," *arXiv preprint arXiv:1301.0567*, 2012.

[9] M. Ponsen, P. Spronck, and K. Tuyls, "Hierarchical reinforcement learning with deictic representation in a computer game," in *Proceedings of the 18th Belgium-Netherlands Conference on Artificial Intelligence*. Citeseer, 2006, pp. 251–258.

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[11] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International conference on machine learning*. PMLR, 2015, pp. 1889–1897.

[12] M. Johnson, K. Hofmann, T. Hutton, and D. Bignell, "The malmo platform for artificial intelligence experimentation." in *IJCAI*. Citeseer, 2016, pp. 4246–4247.

[13] W. H. Guss, B. Houghton, N. Topin, P. Wang, C. Codel, M. Veloso, and R. Salakhutdinov, "Minerl: A large-scale dataset of minecraft demonstrations," *arXiv preprint arXiv:1907.13440*, 2019.

[14] C. Tessler, S. Givony, T. Zahavy, D. Mankowitz, and S. Mannor, "A deep hierarchical approach to lifelong learning in minecraft," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

Acceleration of AI on the edge

[1]*Ladislav POMŠÁR (2nd year),*
Supervisor: [2]*Iveta ZOLOTOVÁ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]ladislav.pomsar@tuke.sk, [2]iveta.zolotova@tuke.sk

***Abstract*—With the emergence of the Internet of Things, the number of devices and data they generate are exponentially growing. In order to handle this growth, the computations traditionally done in the cloud are shifting towards the edge of the network. However, the computational resources are constrained and may not be sufficient to execute advanced artificial intelligence algorithms. In this paper, we provide a brief summary of techniques used to accelerate edge artificial intelligence, summarize our work in the area in the last year, and propose the next steps in our research.**

***Keywords*—edge computing, Edge AI, hardware acceleration, Edge AI accelerators**

I. Introduction

Last year, at this conference, we introduced a review of techniques utilized in the segmentation of volumetric ultrasound, focusing on convolutional neural networks (CNNs). However, due to the coronavirus situation and personal changes in Siemens Healthineers, we decided to retire this topic and introduce the topic of Edge artificial intelligence (AI) acceleration instead.

With an ever-increasing number of connected devices, the need for quick data processing is more urgent than ever. While historically, cloud computing was usually utilized to process these data in data centers, there are three important reasons [1][2] why it is not possible anymore:

- Latency - For real-time decisions, sub-second level responses are necessary
- Data volume and velocity - Generally, physical internet infrastructure is not ready to meet such bandwidth requirements

A paradigm called edge computing [3][4] that shifts computations from the cloud to the devices where data are created have been proposed to mitigate this problem. While this allows for quick data processing, with minimal internet utilization, another problem has

II. Accelerating Edge AI

In order to accelerate Edge AI, several different approaches have been proposed. Different reviews of the state-of-the-art suggest different approaches to the classification. Review [5] lists:

- Model Design - Tries do design shallower neural networks with fewer parameters while keeping accuracy of the deeper and computationally heavier counterparts. This is, however, far from trivial
- Model Compression - This category includes trying to remove unnecessary parameters (pruning), changing representation of parameters (quantization), changing representation of tensors (tensor decomposition), and extracting knowledge from deeper networks (knowledge distillation)
- Inference Acceleration
  - Hardware Acceleration - built-in hardware acceleration is utilized to accelerate neural networks
  - Software Acceleration - an umbrella term for techniques such as model partition, early exit, edge caching, or item filtering

In our research work, we focus on hardware-based inference acceleration in combination with model compression techniques. Hardware acceleration is mostly facilitated via standalone hardware devices, called AI accelerators. As our work is concerned with deploying AI models alongside these devices to the edge of the network, we focus on commercially ready devices (acquirable on the market) with reasonable price, form factor, and power draw that equip specialized AI chips or optimized GPUs. For our work, we call such devices Edge AI Accelerators. For more information, please see author's dissertation proposal or works [5][6][7].

III. Work completed during the last year

A. Making sense of Edge AI accelerators

Edge AI accelerators are mostly thought of as mere tools to launch trained AI algorithms. An algorithm is mostly evaluated on a single Edge AI accelerator to prove that the algorithm can run on the edge in real-time conditions, such as work [8]. Sometimes, several Edge AI accelerators are compared [9] to find the best combination of model and hardware for a task. While this is an understandable position, we think it is not the right one. To maximize the potential of the accelerator, it is necessary to understand its capabilities and constraints fully. To the best of our knowledge, a big survey that would map and describe the capabilities, strengths, and weaknesses of different Edge AI accelerators was not done yet. So far, we have worked on evaluating three different Edge AI accelerators - Jetson Nano, Google Coral USB Accelerator, and Intel Neural Compute Stick 2.

We have also surveyed all commercially available Edge AI Accelerators and are currently working on a publication.

B. Utilizing Edge AI accelerators

A significant part of our research is also built around deploying AI algorithms in the domains of healthcare and

Fig. 1. Jetson Nano utilized in our study.

Industry 4.0.

For Tatra Banka's foundation project, we proposed an application for interpretation of DaT-SPECT, Dopamine Transporter Scan combined with single-photon emission computed tomography, images for detection of Parkinson's disease. Powered by 3D CNN, the system can achieve an accuracy of over 90 percent. We are currently working on deploying this model to an edge device, as the PC on a clinic may not be strong enough to support model runtime in a reasonable time. Other work is concerned with applying different optimization schemes to create a smaller model. Last but not least, we are trying to gather data from a clinic in Kosice to validate the model trained on the publicly available dataset (PPMI) against the real data.

When creating a system utilizing AI, it is typical to create and train your own models. It is also possible to sometimes utilize already existing models. While most Edge AI Accelerators have dedicated optimization frameworks that perform translation of layers from supported framework to the internal framework of the device, quantization, and other steps necessary to run the model on the device, some operate differently via GPU acceleration. We exploited this Nvidia Jetson's (pictured on Figure 1) capability in work [10], where we utilized a pre-trained model from the dlib framework for face detection and recognition. We have created a proof of concept of face recognition powered class attendance system.

IV. Future work

A. Evaluating Edge AI accelerators

This is an extension of our longer effort highlighted in subsection III-A. While we have some preliminary results, it is necessary to evaluate them more in-depth - use a higher number of models, evaluate the effects of quantization on the accuracy or execution speed. Another consideration is that these are just three of a bigger pallet of available devices. It is necessary to expand our hardware portfolio in order to make the comparison as wide as possible. For this purpose, we proposed an FEI Grant project called ACNE - Accelerating computations on the network's edge.

B. Create a guideline for the usage of Edge AI accelerators

The next logical step after evaluating and understanding the Edge AI accelerators is to try to create guidelines for usage of these accelerators. We expect these guidelines to allow for navigation through the existing Edge AI accelerators. However, we would also like to be able to evaluate future accelerators. In order to do so, we already surveyed existing benchmarks of accelerators and planned to evaluate their interpretability for the real-life task. If we will not be satisfied, we will probably resort to implementing our own.

C. Intelligent gateway and federated learning

While it is possible to train an AI model on an edge device, the device's resource budget is usually constrained. Therefore the training is slow and allows to utilize only relatively small models. To mitigate this problem, we would like to explore Edge AI accelerator utilization for federated learning, where several devices are training local models and are then averaged to create a global model. We expect that utilizing local Edge AI accelerators as training devices and local intelligent gateway as an aggregator to accelerate local model training could generate exciting results. We also believe that an intelligent gateway built on the Edge AI accelerator could improve the AI capabilities of the edge. This work will be a continuation of our efforts to improve intelligent gateways, highlighted in [11] we published in 2020.

Acknowledgment

This publication was supported by the grant KEGA – AICybS - Smart Industry/Architectures of Intelligent Information and Cybernetic Systems, 033TUKE-4/2018, 2018-2020 (100 %).

References

[1] B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, and D. S. Nikolopoulos, "Challenges and opportunities in edge computing," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*. IEEE, 2016, pp. 20–26.

[2] J. Pan and J. McElhannon, "Future edge cloud and edge computing for internet of things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2017.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[4] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78–81, 2016.

[5] D. Xu, T. Li, Y. Li, X. Su, S. Tarkoma, T. Jiang, J. Crowcroft, and P. Hui, "Edge intelligence: Architectures, challenges, and applications," *arXiv e-prints*, pp. arXiv–2003, 2020.

[6] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.

[7] L. Deng, G. Li, S. Han, L. Shi, and Y. Xie, "Model compression and hardware acceleration for neural networks: A comprehensive survey," *Proceedings of the IEEE*, vol. 108, no. 4, pp. 485–532, 2020.

[8] M. A. Guillén, A. Llanes, B. Imbernón, R. Martinez-Espana, A. Bueno-Crespo, J.-C. Cano, and J. M. Cecilia, "Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning," *The Journal of Supercomputing*, vol. 77, no. 1, pp. 818–840, 2021.

[9] V. Mazzia, A. Khaliq, F. Salvetti, and M. Chiaberge, "Real-time apple detection system using embedded systems with hardware accelerators: An edge ai application," *IEEE Access*, vol. 8, pp. 9102–9114, 2020.

[10] L. Pomsar, E. Kajati, and I. Zolotova, "Deep learning powered class attendance system based on edge computing," in *ICETA 2020 : 18th IEEE International conference on emerging elearning technologies and applications*, November 2020.

[11] M. Kvetko, J. Mocnej, L. Pomšár, and I. Zolotová, "Raspberry Pi And Windows 10 powered intelligent modular gateway for decentralized IoT environments," *Acta Electrotechnica et Informatica*, vol. 20, no. 2, pp. 44–50, Jun. 2020. [Online]. Available: https://doi.org/10.15546/aeei-2020-0012

Study of visualization methods of explainable artificial intelligence

[1]*Ivan Čík (1st year),*
Supervisor: [2]*Marián MACH*

[1,2]Dept. Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]ivan.cik@tuke.sk, [2]marian.mach@tuke.sk

***Abstract*—Nowadays, artificial intelligence is a trend that is used in every application field,e.g. in finance, healthcare, sport, music, military, and many others. The complexity of the models has led to certain questions about their explainability. Uncertainty and lack of confidence in model decisions provided the ground for the emergence of an area of explainable artificial intelligence. The article describes the basic XAI visualization methods (LRP, Grad-CAM) and experiments with three datasets (Chicago Face Database, MNIST Database, MNIST-Fashion dataset).**

***Keywords*—Explainable Artificial Intelligence, LRP, Grad-CAM, CNN**

I. Introduction

The increase in computational power has allowed the emergence of a sub-area of artificial intelligence called deep learning and we also achieved super performance in many tasks that were previously thought to be computationally unattainable [1]. Machine learning models for solving complex tasks often have several million parameters, which becomes incomprehensible to the human user. Lack of transparency and interpretability is the issue with many state-of-the-art models. In some cases, it would be unbearable for us to know what is happening in these algorithms, such as medical diagnosis, where a confidence criterion is a basis for a model decision. The key solution would be to know why the algorithm output is as trustworthy as obtaining this information would not reduce the performance of the algorithm. Due to these problems, **explainable artificial intelligence (XAI)** has become an area of interest in the research community. This article describes visualization methods that explain through backpropagation: **Layer-Wise Relevance (LRP [2]), Gradient-weighted Class Activation Mapping(Grad-CAM [3])**.

II. Visualization back-propagation based methods

The output from the XAI visualization method is the identification of areas (pixels at the input, features, or areas) from the input that significantly affect the outputs of the model (*e.g.,* saliency maps). The principle is to visualize the relevance/saliency of a feature based on the volume gradient that has passed across the layers of neural networks during training (evaluated from output to input). The basic approach is to visualize the partial derivation of the network output w.r.t. each input feature (*e.g.,*, the task of recognizing activity, particularly in football, may have high relevance to pixels representing the ball and CNN players, which may mean that CNN's decision is sensitive to the presence of the ball in the image) [4].

A. Grad-CAM

The gradient-based interpretability method uses the reverse transmission of information in the neural network to understand the influence of neurons and the correlation between input x and output y. Gradient-weighted Class Activation Mapping(Grad-CAM) method [3] is a class-discriminative attribution technique used to distinguish the neuronal activity of CNN network neurons. Grad-CAM uses the gradients of any class-specific image that flows into the final convolutional layer to generate a map, highlighting the important areas in the image which contributed to particular model output.

B. LRP

LRP method explains deep neural network prediction and is one of the relevance propagation methods so that the prediction propagates backward without the use of gradients. The algorithm starts from the output layer L of the deep neural network, and then moves in the opposite direction in the graph, gradually redistributing the prediction score until the input is reached. Each neuron receives a share in the output of the network and further redistributes it to its predecessors in the same amount until the input variables are reached [5], [6]. LRP is a method of identifying important pixels by inverting a neural network. The reverse pass is a conservative redistribution process, in which the neurons that contribute the most to the previous layer are the most important. Mathematically, it redistributes the prediction $f(x)$ backward, using redistribution rules, until it assigns a relevance score R_i to each variable. An important feature of the redistribution process is called global conservation property and is given as:

$$\sum_i R_i = ... = \sum_j R_j = \sum_k R_k = ... = f(x) \qquad (1)$$

The described property says that in each step of the redistribution process the overall relevance is maintained.

III. Convolutional Neural Networks

Convolutional neural networks have achieved tremendous success on many different learning tasks (in some they have performed poorly than humans, but there are also cases where they have achieved the same or even better performance), including image recognition [7], [8]. Compared to other computer vision methods, they can be trained fully autonomously with only little pre-processing. The ability to allocate computational processes to a graphics processing

unit has accelerated the training process and brought an explosion of use due to a speed-up in training time through parallelization. Convolutional Neural Network architectures are providing good performance on various tasks such as object detection, face recognition, and more. In [9], concluded that with careful scaling of the model (model depth, width, and resolution) we can effectively fine-tune the model to achieve better performance.

IV. Experiments

A. Grad-CAM Experiments

A group of scientists in the field of psychology and the study of emotions has compiled the Chicago Face Database [10]. It is a freely available database consisting of high-resolution, standardized photos of 158 subjects of various ethnicity aged 17 to 65 years, expressing different emotions. Each face in the database had one expression among five: **neutral, angry, fearful, happy with mouth closed, happy with mouth opened**. The neural network that was used in experiments in itself was made of 5 layers. The first 3 layers are intertwined convolution layers and batch normalization layers. The convolution layers let the network pick out features from the picture, and each batch normalization keeps the values in ranges that let the network reach useful conclusions followed by a concatenate pooling layers: one is a maximum pooling, the second an average pooling. Finally, the last 2 layers are fully dense layers used for the classification of features learned by the convolutions. For the training of the network, a learning rate of 0.001 was chosen after some trial and error. Adam [11] was used as optimizer. The final accuracy acquired on the validation dataset was of around 65% across 5 different run. To add a measure of understanding to the artificial neural network, we decided to add a visualization of the areas that hold high importance in the decision process of the neural network. To accomplish that, we relied on the Gradient-weighted Class Activation Mapping(Grad-CAM).

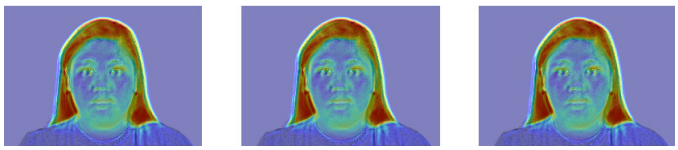


Fig. 1. Output of Grad-CAM on facial expression with Neutral Emotion. Note that the pixels that were crucial in recognizing the emotion are mostly allocated to the hair, and we would rather expect it to be on the face.

The GRAD-CAM method did not add much insight into the process of the artificial neural network either. The method that would display this information as heat maps over the images. The result of Grad-CAM (see Figure 1) of that endeavor is something that we were not expecting: the heat-map displays the hair as being the highest importance for the network, with the facial features coming after in intensity. As an output, we would expect the class-discriminative localization map to be located mainly in the facial area, as we trained our network to recognize emotions using facial expression.

B. LRP Experiments

Our goal was to interpret the outputs from neural networks. We tried to identify which areas were crucial for classification at the input. Subsequently, we tried to create a cumulative map for the whole group of inputs (class) and not just for one specific input. We have used the specific ϵ-LRP method. In our experiments, we used the well-known MNIST dataset [12], which consists of handwritten numbers from 0 to 9, so it contains 10 classes. Images measure 28 x 28 x 1 pixels. The dataset contains 60,000 training and 10,000 test images. The neural network model had 2 hidden layers, each with 256 neurons. On testing dataset, we achieved an accuracy of 96.75%.

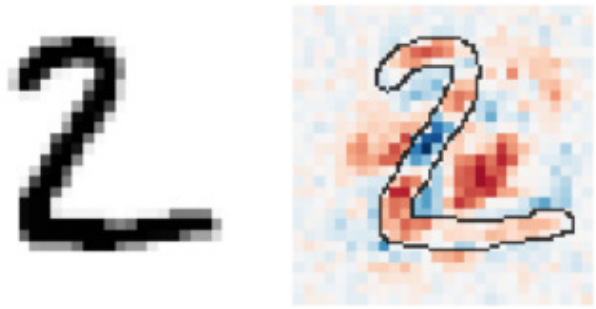


Fig. 2. Visualization of output and heatmap by ϵ-LRP method. The value of red pixels is positive, or blue if their value is negative. The intensity of the color depends on the distance of value from zero. Dark red pixels had the greatest influence on the neural network decision when choosing a class. [13]

For further experiments with this method, we decided to use a more complex dataset. We used the MNIST-Fashion dataset[14], which has 10 classes, each class including a certain type of clothing (T-Shirt/Top, Trouser, Pullover, Dress, Coat, Sandals, Shirt, Sneaker, Bag, and Ankle boots). The amount of data, architecture, number of epochs, and all parameter settings for training remained the same as in the first experiment, but the accuracy was 86.14%, which was caused by the higher complexity of the MNIST-fashion dataset.

First, we took all the test images that belong to the "Coat" class. For each image in this class, we used the LRP method and added the individual values at the same positions, and then divided them by the number of images.

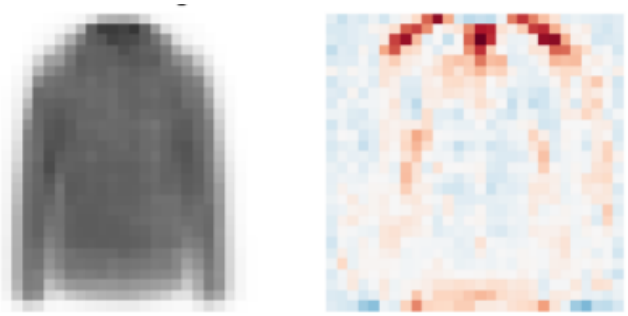


Fig. 3. Cumulative heatmap of the "Coats" class.

The output is the average heatmap for a given class, which should visualize which area of the image most influences the classification decision for a given class in general. The problem of misclassification comes when we have similar heatmaps for two classes, e.g. dog breed. If our accuracy is not high enough, we would look at classes that have very similar cumulative heatmaps, and in the training pictures, we would try to cover the areas on which they learn. In this way, we could get our model to consider other areas than identifying the class.

V. Conclusion and Future Plan

Analysis showed that it did so by focusing on parts of the images that are not so important. This mainly shows that our networks try to identify patterns that are not important. Another hypothesis and more probable is that the network identified and learned a bias present in the data that the annotator nor we humans did not identify. Such bias generally harms the performance of Deep Neural Networks when used in a real-world application. Therefore choosing the right dataset and transformation is useful for DNN networks to properly learn the foreground pattern. In future works, we will investigate models trained on wild datasets. Our goal is to increase the generalization of CNN so that the classification is not sensitive to random features that do not actually describe the object. Here are two scientific goals we want to address in further research:

- Introduction of a metric for measuring the quality of the CNN model. By comparing the class activation map (evaluated on different visualization XAI approaches, see Figure 1) and segmentation of an object from an image, we can identify whether the model focuses on the correct parts of the image. By using the proper loss function, we should be able to maintain the model in the areas that matter by punishing neurons that should not be activated highly. This can lead to improvement of the developer's understanding of CNN model behavior and potentially increasing CNN model performance in terms of accuracy and resilience to background change.
- Introduction of similarity or difference of data points to other classes, what can be the source for later explanation. By obtaining a description of data points based on the calculation of the distance between data points and centroids of other clusters or cluster edges before the classification layer in the model. Let's have a trained CNN model f for object recognition. By removing this fully connected layer and then applying clustering, we would like to split the individual inputs into clusters. The number of clusters is equal to the number of classes of recognized objects. Subsequently, we would like to calculate the distance to the centroids of another class or to the boundaries of the second cluster for a given input and thus introduce a similarity metric. We would also use these distance values with the input image as input data for the classification part of the model. Decision support and possible acquisition of new knowledge about data. Increase end-user trust in model outputs. Promoting an understanding of how the model works by describing the distance between objects. Gaining new knowledge and insights into the model. Improving the explanation.

Acknowledgment

This research is supported by AI4EU project from the European Union's Horizon 2020 research and innovation program under grant agreement 825619 2019-2021, and the Maria Currie RISE LIFEBOTS Exchange Grant Agreement ID : 824047 2019-2021, and FEIGRANT – 2021 – Increase server computing power to work with state-of-the-art artificial intelligence algorithms.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] S. Lapuschkin, A. Binder, G. Montavon, K.-R. Muller, and W. Samek, "Analyzing classifiers: Fisher vectors and deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2912–2920.

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[4] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[5] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.

[6] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 11, pp. 2660–2673, 2016.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[8] N. J. Nilsson, "Stuart russell and peter norvig, artificial intelligence: A modern approach," *Artificial intelligence*, vol. 82, no. 1-2, pp. 369–380, 1996.

[9] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[10] D. S. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behavior research methods*, vol. 47, no. 4, pp. 1122–1135, 2015.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[13] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, "Layer-wise relevance propagation: an overview," in *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer, 2019, pp. 193–209.

[14] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

Review of Multiport Converters Topologies

[1]*Adrián MARCINEK (1st year)*
Supervisor: [2]František ĎUROVSKÝ

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

[1]adrian.marcinek@tuke.sk, [2]frantisek.durovsky@tuke.sk

***Abstract*— The article provides a comprehensive view of the topologies used to create multiport converters. The main advantages of multiport converters are their ability to integrate different energy sources and to adapt the characteristics of the consumed or supplied energy to these sources. By characteristic is meant, for example, MPPT (Maximum Power Point Tracking) for PV (photovoltaic), or battery charging characteristic.**

***Keywords*—renewable energy source, hybrid vehicle, electric vehicle, power converter, multiport power converter**

I. INTRODUCTION

Nowadays, we are increasingly confronted with a trend of reducing emissions, especially CO_2. These emissions are produced by cars, trucks, ships, or aircrafts. Internal combustion engines belong to the largest producers of emissions in transport. Emissions are also produced by big electricity producers such as coal-fired power plants. There are several ways how CO_2 emissions can be reduced from transport. These ways are represented by hybrid, plug-in hybrid, and electric vehicles, which make a major contribution to reducing CO_2 emissions. Emissions reductions can also be achieved in coal-fired power plants. They are being replaced by renewable energy sources. The most common renewable energy sources are photovoltaic panels, wind turbines, fuel cells, etc.

All types of energy converters in these devices have one thing in common - DC/DC or DC/AC power converters, which connect multiple power sources to multiple loads. These converters connect sources and loads with different voltage levels. To make a system powered by a renewable source useful, it is necessary to ensure the lowest possible losses in the transfer of energy between sources and loads and as few conversion steps as possible. Most frequently used topology uses one common bus with a defined voltage and several DC/DC or DC/AC converters. However, this topology has number of disadvantages, for example many steps of energy conversions, which reduces the efficiency of the system. Using multiple DC/DC or DC/AC converters, the dimensions of all system, as well as financial costs, increase and power density decrease. To improve the overall efficiency of the system and reduce the number of energy conversions, systems with a common bus began to be replaced by multiport power converters. In part II. the topologies of converters used in renewable energy sources will be described. In part III. the topologies used in multiport converters will be described and part IV will summarize the findings from the described topologies and the direction of further research.

II. POWER CONVERTERS FOR RENEWABLE ENERGY SOURCES

Many different power converter topologies, which include PV (photovoltaic) and ES (Energy Storage) systems, were introduced in [1] – [5]. Fig. 1 shows four dominant power converter topologies used today for renewable energy sources. The most widely used topology includes separate DC/DC converters for RES (Renewable Energy Source) and ES (Energy Storage) systems, which convert power to a common DC bus. This topology is shown in Fig. 1 (a). The DC/DC converter that control charging and discharging the battery must be bidirectional. In the next step, the energy from the DC bus is transformed by a DC/AC power converter to AC line. The second most popular topology is the topology with an integrated battery on the DC bus, which is demonstrated in Fig. 1 (b). In this way, the simplification of the whole system can be achieved by reducing the number of power converters. Another topology (Fig. 1 (c)) uses the same number of power converters as the first mentioned topology, but the battery is connected on the AC line side through DC/AC converter. Thus, this topology contains two DC/AC converters, while the DC/AC converter for the battery is bidirectional. Each of the mentioned topologies has its advantages as well as disadvantages. In comparison, we will consider a system that has a renewable energy source at the input, represented by a photovoltaic panel, and a battery will serve as energy storage. This system will be further connected to the load and to the AC line using a DC/AC converter. The topology in Fig. 1 (a) is suitable for batteries with lower voltage, where the DC/DC converter for the battery adjusts the voltage level of the battery to voltage level of DC bus. For the topologies shown in Fig. 1 (b) and Fig. 1 (c) it is better to use a battery with a higher voltage, due to voltage compatibility with the AC line.

In practice, for the topology in Fig. 1 (b) it is necessary to add a relays or switches that will control the flow of energy from PV to the battery or from PV to the AC line. Because the battery in this topology is connected directly to the DC bus, depending on the state of charge (SOC) of the battery and the DC bus voltage, the DC/AC converter must be controlled to maintain synchronization with the AC line. The converter topology shown in Fig. 1 (d) solves many of the shortcomings of the converters mentioned so far. This system includes a multiport converter and is much more flexible. Topology uses fewer converters to achieve the same performance as the topologies before. Thus, the multiport converter will ensure a higher power density of the entire system and increase its efficiency. The multiport converter, unlike previous simple

topologies, requires fewer conversion steps to achieve the desired voltage conversion.

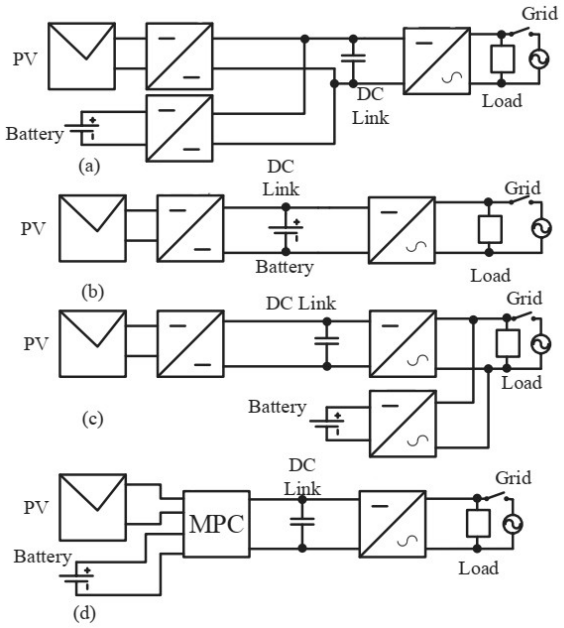


Fig. 1. Basic topologies for photovoltaic and energy storage, [1], [2], [5]-[27],
(a) Multi converter topology
(b) DC link – battery topology
(c) AC side – battery topology
(d) Integrated three port converter

Of course, such a multiport converter is more difficult to control, but the overall flexibility of the operation of the power converter increases. The main goal of multiport converters is to integrate several ports into one power converter, which will allow energy to flow between the connected devices.

III. MULTIPORT CONVERTER TOPOLOGIES

Several different topologies of power converters were examined for comparison. All examined topologies can be divided into two main categories: "non-isolated" and "isolated" power converters. Based on the name, isolated multiport converter topologies use a transformer with a suitable transformation ratio and buck or boost converter. Isolated topologies generally offer a greater voltage range and flexibility of power flow between ports. Non-isolated topologies are more efficient because they do not have any transformer losses.

Non-isolated topologies can be divided into three groups:
1. Combined input/output ports
2. Magnetic and Capacitive coupled ports
3. Reconfigurable ports

The topology of the combined input/output ports multiport converters is based on a combination of various basic converters such as a buck converter, a boost converter, or a buck - boost converter. Moreover, complex converters consist of a half-bridge or full bridge of transistors. Topologies with reconfigurable ports are simpler in construction. This topology usually uses one or two converters to connect multiple ports between each other. These converters are further supplemented by several relays or other switching systems for rearranging the topology. Thus, same power converter can exchange energy between other ports. The topology of the converter using inductive or capacitive coupling is based on a combination of inductance or capacitor with the basic converters. Magnetic or capacitive coupling does not mean galvanic isolation, but elements such as inductance and capacitor are used in the converter to increase the voltage. This is useful for renewable energy sources and battery, where a wider voltage range can be used.

Like non-isolated topologies, also isolated topologies can be divided into three groups:
1. Two winding transformer coupled ports
2. Multi winding transformer coupled ports
3. Multi transformer coupled ports

The first of these topologies uses a single-core, two-winding transformer, where usually one or more ports are combined on the primary or secondary side of the transformer. The second topology is used, when three or more ports are connected by one multi-winding high-frequency transformer, which serves as an isolation for all ports. The third topology uses several transformers as the basis of an isolated multiport converter. In the following subchapters, the individual topologies of non-isolated and isolated converters will be discussed in detail.

A. Non-isolated topologies

1) Combined input/output ports: In [7], the authors describe a family of non-isolated converters based on buck, boost, and buck - boost converters in various combinations. However, the use of such converters allows the flow of energy only in one direction, respectively to load port. In [6] was showed charging and discharging to and from ports of converter based on the use of a bidirectional converter. In [7] it was stated that a multiport converter could be created using a suitable combination of basic converters. In [8], the authors present a converter with a cascade boost converter and a bidirectional buck converter to create a bidirectional power flow between the ports, where the energy storage port is one of the inputs and the load is powered from a common DC bus. In [9], a very simple multiport converter is introduced. Different power sources can be connected to each of its input ports. This is achieved by parallel combination of several buck - boost or SEPIC converters. However, such a multi-input solution does not transfer energy from several sources to the load, because a circulating current would be generated in circuit. Because this topology is based on a combination of basic converters, not all ports will participate in the power flow at the same time. Most of these topologies, where basic converters are combined, primarily transfer energy between two ports, or between two ports and a third port (dual-input or dual-output). For dual input or dual output, the coil current is the sum of the currents from the two inputs. Depending on the control, one port supplies power only for a certain time compared to the other port. This is mainly used in PV-battery systems, where the battery port is not continuously active. However, these topologies also have limitations. It is necessary to block the return flow of energy to the PV system and this blocking is usually solved by a diode, which, however, causes losses

during normal operation. In [10], a topology is presented, where multiple power sources are connected by coils to individual half-bridges, which are further connected to a common DC bus. Then, the output AC port is powered by the DC bus using an inverter and a low-frequency transformer. The AC port of the converter is isolated by a transformer from all connected sources. However, the renewable energy sources not being isolated from each other could cause circuit currents between these sources if they worked in parallel on one DC bus. In [11] a multiport converter with the possibility of connection to an AC line is presented. Its main advantage is that it converts energy from the PV panel and battery directly into the AC line without isolation between them. The coil or capacitor between the DC/DC and DC/AC converters act as a connection to the AC line. This allows us to reduce the number of steps of conversion. However, the number of active elements of the converter is increased, which decreases the efficiency of the system.

2) Magnetic and capacitive coupled ports: The authors in [12] report non-isolated multiport power converter, which is a combination of boost converters. Commonly used coils are replaced by coupled coils for higher voltage gain. This topology is especially useful for renewable sources with a wide voltage range. Active clamp circuit allows soft switching and increases performance. In [13], a similar topology is presented, where energy from the input ports is transferred to the output port by means of a boost converter. A sufficient increase in the voltage of the output port is caused by using coupled coils. Capacitive coupling is given in [14], where multi-input ports are combined with capacitive coupling in two steps. Capacitive coupling is based on the principle of a charge pump and an increase output voltage is based on the number of step-up steps with diodes and capacitors.

3) Reconfigurable ports: In [15] the three-phase inverter is reconfigured into an AC / DC converter to charge the battery from the AC line and to charge the battery from the PV panels using a three-phase bridge. This topology provides few conversion steps and a small number of elements. However, the flexibility of these topologies is small because reconfiguring the converter to work in other modes is complex and requires many switches. Similarly, in [16] and [17], the converters can be configured to flow power between multiple ports using slow switches such as relays. In [17], three phases from a three-phase bridge are converted to one phase using a common mode circuit. In this case, it is possible to use less switches or relays and increase the efficiency of the system. However, the AC output is only single-phase.

B. Isolated topologies

1) Two winding transformer coupled ports: In [18] and [19] the authors present topologies that integrate PV, battery and load port based on the principle of active clamped forward converter. According to [20], for PV applications, two of the three converter ports must be controlled independently, switches S1 and S2 must also be controlled independently to realize port 1 to port 2 power flow. This requirement is achieved in Fig. 2 (a) by means of diode D1 and switch S3 and in Fig. 2 (b) by shorting the switches S3 and S4 on the secondary side of transformer.

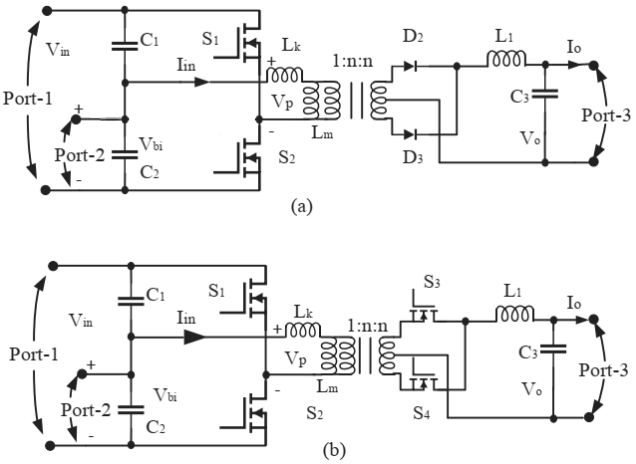


Fig. 2. Isolated topology with two ports, [18], [20],
(a) Primary freewheel
(b) Secondary freewheel

After that, the bidirectional port allows current to flow from and to the port, while switches S1 or S2 are on. When switch S3 is on, the current is zero. In this way, it is possible to influence the output power over time. Although, power conversion between the three ports is possible, the flow of power to and from the bidirectional port is difficult to implement. The topology given in [21] consists of ports connected in cascade at the input of the bridge converter on the primary side of the transformer. In [22], the third port was combined using boost converter on the secondary side of the transformer of the half-bridge converter. Topology with increasing voltage on the primary side of the transformer was presented in [23] and [24]. The topology in [24] uses fewer active elements than the topology in [23], so converter is cheaper. However, the topology in [23] allows better control options. In [25], the author describes two dual boost converters coupled to transformers with two windings. There are four different ports available in this topology, all of them can be bidirectional. The main advantage of this topology is its flexibility.

2) Multi-winding transformer coupled ports: Another type of isolated multiport converters are converters using multi-winding transformers. These transformers provide isolated power transfer between all ports and typically include half-bridge [26] or full bridge [27] DC/AC converter topology. These topologies allow bidirectional power flow between ports. The topology is shown in Fig. 3. In [27], a resonant circuit was also included in the multiport converter, which allowed to increase the switching frequency and reduced losses. In published topologies that uses only half-bridge converter only 50% of the duty cycle can be achieved during operation. Soft switching in this topology is also difficult to achieve.

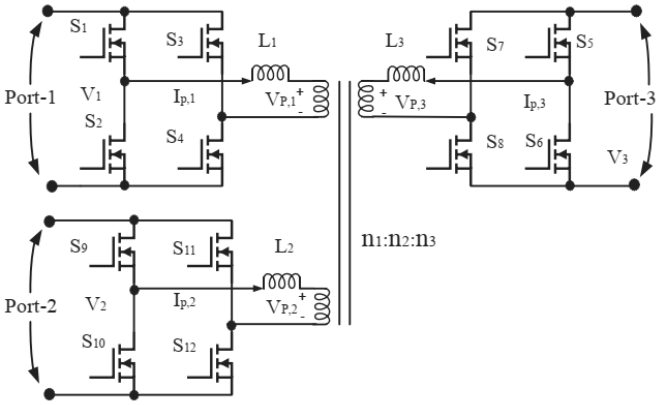


Fig. 3. Isolated triple active full bridge converter [25]-[27]

3) Multi-transformer coupled ports: The last topology of isolated multiport converters is the topology given in [28] and [29]. This topology, as before, provides isolation of all ports of the multiport converter and allows power to flow from and to multiple ports. The two transformers contain active bridges on their primary side. The secondary windings of these two transformers are connected by means of a three-phase bridge and form the third port of the multiport converter. The topology is shown in Fig. 4. The multiport converter with this topology can be controlled by a PWM signal that is suitably phase shifted for each half-bridge. This topology is most advantageous in terms of multiport converter modularity. Modularity can include changing the number of ports, changing the power of one of the ports, and changing the type of source or load connected to the converter. These changes should be as simple as possible. One way how to modify multiport converter is to modify its transformer as it is described in [30]. The authors present the possibilities of adding serial or parallel windings to already existing transformers using modular cores.

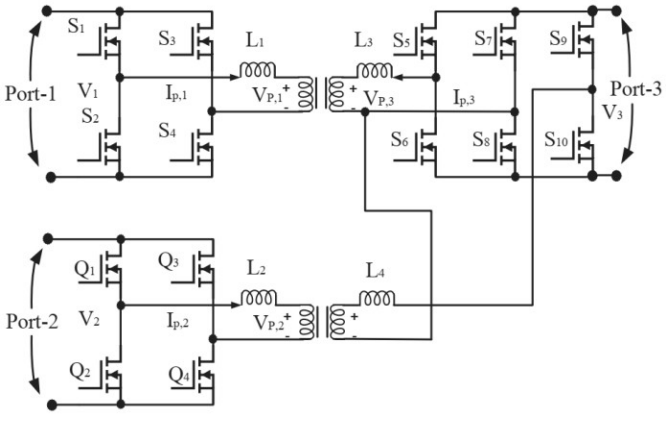


Fig. 4. Triple port active bridge converter based on dual transformer [28] and [29]

IV. Conclusion

Multiport converters have an advantage over conventional structures mainly in the number of used elements in converters and in the number of conversion steps. Also, losses are lower, because these converters combine and share some elements, such as high-frequency transformers between the converter ports. However, multiport converters are more complex and therefore their control is more difficult. In our further research, we will focus on isolated multiport converters with a high degree of modularity. This means that topologies, where a design intervention in the transformer is required, are not suitable for us. These topologies include converters with multi-winding transformers that are designed for a specific application. Although, the removal of the winding from the transformer core is not complicated, the resulting air gap will reduce the efficiency of the transformer. By adding additional windings, we can encounter a design limitation of the transformer core. Therefore, when the number of ports in a such multiport converter is changed, the transformer should be replaced. In our further research, we will focus exclusively on multiport converters with several double-winding transformers with secondary windings connected in parallel. In addition to high modularity, low weight must be guaranteed, which can be achieved by using the planar transformers with low losses and high heat dissipation. In the multiport converter designed in this way, we will focus mainly on the possibilities of its control.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0750.

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-18-0436.

References

[1] H. Mahmood, D. Michaelson and J. Jiang, "A Power Management Strategy for PV/Battery Hybrid Systems in Islanded Microgrids," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 2, no. 4, pp. 870-882, Dec. 2014.

[2] B. Wang, M. Sechilariu and F. Locment, "Intelligent DC Microgrid With Smart Grid Communications: Control Strategy Consideration and Design," in IEEE Transactions on Smart Grid, vol. 3, no. 4, pp. 2148-2156, Dec. 2012

[3] B. Wang, M. Sechilariu and F. Locment, "Intelligent DC Microgrid With Smart Grid Communications: Control Strategy Consideration and Design," in IEEE Transactions on Smart Grid, vol. 3, no. 4, pp. 2148-2156, Dec. 2012

[4] K. Sun, L. Zhang, Y. Xing and J. M. Guerrero, "A Distributed Control Strategy Based on DC Bus Signaling for Modular Photovoltaic Generation Systems With Battery Energy Storage," in IEEE Transactions on Power Electronics, vol. 26, no. 10, pp. 3032-3045, Oct. 2011.

[5] H. Zhou, T. Bhattacharya, D. Tran, T. S. T. Siew and A. M. Khambadkone, "Composite Energy Storage System Involving Battery and Ultracapacitor With Dynamic Energy Management in Microgrid Applications," in IEEE Transactions on Power Electronics, vol. 26, no. 3, pp. 923-930, March 2011.

[6] Meng Zhang, Y. Xing, H. Wu, Yangjun Lu and Kai Sun, "Performance evaluation of a non-isolated bidirectional three-port power converter for energy storage applications," 2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia), Hefei, 2016, pp. 2703-2708.

[7] Hongfei Wu, Yan Xing, Yanbing Xia and Kai Sun, "A family of non-isolated three-port converters for stand-alone renewable power system," IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society, Melbourne, VIC, 2011, pp. 1030-1035.

[8] H. Zhu, D. Zhang, B. Zhang and Z. Zhou, "A Nonisolated Three-Port DCDC Converter and Three-Domain Control Method for PV-Battery Power Systems," in IEEE Transactions on Industrial Electronics, vol. 62, no. 8, pp. 4937-4947, Aug. 2015.

[9] A. Kwasinski, "Identification of Feasible Topologies for Multiple-Input DC-DC Converters," in IEEE Transactions on Power Electronics, vol. 24, no. 3, pp. 856-861, March 2009.

[10] M. McDonough, "Integration of Inductively Coupled Power Transfer and Hybrid Energy Storage System: A Multiport Power Electronics Interface for Battery-Powered Electric Vehicles," in IEEE Transactions on Power Electronics, vol. 30, no. 11, pp. 6423-6433, Nov. 2015.

[11] M. Amirabadi, H. A. Toliyat and W. C. Alexander, "A Multiport AC Link PV Inverter With Reduced Size and Weight for Stand-Alone Application," in IEEE Transactions on Industry Applications, vol. 49, no. 5, pp. 2217-2228, Sept.-Oct. 2013.

[12] Y. M. Chen, A. Q. Huang and X. Yu, "A High Step-Up Three-Port DCDC Converter for Stand-Alone PV/Battery Power Systems," in IEEE Transactions on Power Electronics, vol. 28, no. 11, pp. 5049-5062, Nov. 2013.

[13] L. J. Chien, C. C. Chen, J. F. Chen and Y. P. Hsieh, "Novel Three-Port Converter With High-Voltage Gain," in IEEE Transactions on Power Electronics, vol. 29, no. 9, pp. 4693-4703, Sept. 2014.

[14] V. A. K. Prabhala, P. Fajri, V. S. P. Gouribhatla, B. P. Baddipadiga and M. Ferdowsi, "A DC-DC Converter With High Voltage Gain and Two Input Boost Stages," in IEEE Transactions on Power Electronics, vol. 31, no. 6, pp. 4206-4215, June 2016.

[15] H. Kim, B. Parkhideh, T. D. Bongers and H. Gao, "Reconfigurable Solar Converter: A Single-Stage Power Conversion PV-Battery System," in IEEE Transactions on Power Electronics, vol. 28, no. 8, pp. 3788-3797, Aug. 2013.

[16] T. Kim and S. Kwak, "Single pole switch leg based multi-port converter with an energy storage," in IET Power Electronics, vol. 9, no. 6, pp. 1322-1330, 5 18 2016.

[17] J. M. Shen, H. L. Jou and J. C. Wu, "Transformer-less three-port gridconnected power converter for distribution power generation system with dual renewable energy sources," in IET Power Electronics, vol. 5, no. 4, pp. 501-509, April 2012.

[18] Z. Qian, O. Abdel-Rahman, H. Hu and I. Batarseh, "An integrated three-port inverter for stand-alone PV applications," 2010 IEEE Energy Conversion Congress and Exposition, Atlanta, GA, 2010, pp. 1471-1478.

[19] Z. Qian, O. Abdel-Rahman, H. Al-Atrash and I. Batarseh, "Modeling and Control of Three-Port DC/DC Converter Interface for Satellite Applications," in IEEE Transactions on Power Electronics, vol. 25, no. 3, pp. 637-649, March 2010.

[20] H. Wu, R. Chen, J. Zhang, Y. Xing, H. Hu and H. Ge, "A Family of Three-Port Half-Bridge Converters for a Stand-Alone Renewable Power System," in IEEE Transactions on Power Electronics, vol. 26, no. 9, pp. 2697-2706, Sept. 2011.

[21] H. Wu, J. Zhang and Y. Xing, "A Family of Multiport Buck-Boost Converters Based on DC-Link-Inductors (DLIs)," in IEEE Transactions on Power Electronics, vol. 30, no. 2, pp. 735-746, Feb. 2015.

[22] H. Zhu, D. Zhang, H. S. Athab, B. Wu and Y. Gu, "PV Isolated ThreePort Converter and Energy-Balancing Control Method for PV-Battery Power Supply Applications," in IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3595-3606, June 2015.

[23] H. Wu, J. Zhang, X. Qin, T. Mu and Y. Xing, "Secondary-SideRegulated Soft-Switching Full-Bridge Three-Port Converter Based on Bridgeless Boost Rectifier and Bidirectional Converter for Multiple Energy Interface," in IEEE Transactions on Power Electronics, vol. 31, no. 7, pp. 4847-4860, July 2016.

[24] W. Li, J. Xiao, Y. Zhao and X. He, "PWM Plus Phase Angle Shift (PPAS) Control Scheme for Combined Multiport DC/DC Converters," in IEEE Transactions on Power Electronics, vol. 27, no. 3, pp. 1479-1489, March 2012.

[25] Z. Ding, C. Yang, Z. Zhang, C. Wang and S. Xie, "A Novel SoftSwitching Multiport Bidirectional DC-DC Converter for Hybrid Energy Storage System," in IEEE Transactions on Power Electronics, vol. 29, no. 4, pp. 1595-1609, April 2014.

[26] L. Wang, Z. Wang and H. Li, "Asymmetrical Duty Cycle Control and Decoupled Power Flow Design of a Three-port Bidirectional DC-DC Converter for Fuel Cell Vehicle Application," in IEEE Transactions on Power Electronics, vol. 27, no. 2, pp. 891-904, Feb. 2012.

[27] H. Krishnaswami and N. Mohan, "Three-Port Series-Resonant DC-DC Converter to Interface Renewable Energy Sources With Bidirectional Load and Energy Storage Ports," in IEEE Transactions on Power Electronics, vol. 24, no. 10, pp. 2289-2297, Oct. 2009.

[28] E. Asa, K. Colak, M. Bojarski and D. Czarkowski, "Asymmetrical Duty-Cycle and Phase-Shift Control of a Novel Multiport CLL Resonant Converter," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 3, no. 4, pp. 1122-1131, Dec. 2015.

[29] V. N. S. R. Jakka, A. Shukla and G. D. Demetriades, "DualTransformer-Based Asymmetrical Triple-Port Active Bridge (DTATAB) Isolated DC-DC Converter," in IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 4549-4560, June 2017.

[30] Dalton Honorio, Luiz Henrique S. C. Barreto, Andrew Foote, Burak Ozipineci, João Onofre Pereira Pinto, "Modular transformer in isolated multiport power converters" in IEEE Southern Power Electronics Conference (SPEC), Puerto Varas, 4-7 Dec. 2017

Contribution to Generative Enhancement of 3D Image Classifiers

[1]*Michal VARGA (5th year),*
Supervisor: [2]*Ján JADLOVSKÝ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]michal.varga@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract*—This paper summarizes the past year's increment in our research of generative enhancement of 3D image classifiers. Three major contributions are presented: generative enhancement of the VoxNet classifier, evaluation of the training set size effect on the accuracy of classifiers and their generative variants, and evaluation of generative enhancement under real-world conditions.**

***Keywords*—3D image classification, deep learning, convolutional neural networks, generative modeling, 3D imaging**

I. Introduction

Our research focuses on the application of generative modeling in the classification of 3D images. The success of generative classifiers in various public datasets motivated us to investigate this area and the effect of generative modeling on classification performance. Close competition between generative and non-generative classifiers suggests that both approaches have their respective strengths and are both worthy of pursuit. We also realized that under some conditions, these two methods of building classifiers are not mutually exclusive and there exists a potential to combine their strengths. The problems, that the generative approach attempts to solve, are low classification accuracy and generalization ability of existing classifiers. Especially in the domain of 3D classification, where the abundance of training data is not common, the augmentative powers of generative modeling could aid in achieving high accuracy. This motivated us to design the methodology of generative enhancement of 3D image classifiers. The methodology defines the process of synthesis of novel generative classifiers from existing non-generative classifiers by embedding them in a generative architecture.

II. Initial status

The proposed methodology has been published in *MDPI Applied Sciences* [1]. It consists of three steps: designing a compatible generator, modifying the classifier to discriminate fake and real samples, and updating the training loop to contain adversarial optimization. Several experiments evaluating the effects of the methodology have been conducted and published in the article. In the experiments we applied the methodology to a simple 3D convolutional classifier labeled CNN (Convolutional Neural Network) and comparing its original accuracy and precision with the resulting generative classifier labeled CGAN (Conditional Generative Adversarial Network). Output of the trained generator is shown in Fig. 1.

TABLE I: Accuracy and precision of classifiers (ModelNet10)

	Accuracy [%]	mAP [%]	Rel. accuracy [%]
CNN	83.04	83.75	100.00
augCNN	84.58	86.31	101.85
CGAN	**89.21**	**89.18**	**107.43**
VoxNet	85.79	89.10	100.00
augVoxNet	87.00	89.53	101.41
genVoxNet	**87.67**	**90.21**	**102.19**

The results show a 7.43% improvement of accuracy without any adverse computational impact on the inference performance. The augmented variant labeled augCNN (augmented CNN; trained on both training set and generator output), achieved lower (1.85%) performance increase, which suggests the benefit of the generative enhancement is not caused solely by the augmentative effect of the generator.

III. Current Research

In the past year, we examined the effects of the methodology on an already existing 3D image classifier—VoxNet [2]. We also performed experiments to verify the idea, that the generative enhancement can improve the generalization potential of the original classifier. Additionally, a rotary 3D scanner has been designed and constructed to evaluate the viability of the methodology under real-world conditions.

A. Generative enhancement of the VoxNet classifier

VoxNet is a 3D voxel-based convolutional classifer [2]. We applied the proposed methodology by inserting the VoxNet classifier into a GAN-like generative network. Since the format of the input is similar to the classifier in our previous experiment, we could reuse the generator. The classifier itself and the training loop have been adjusted according to the methodology process. The resulting generative modification (genVoxNet) has been evaluated on the ModelNet10 dataset [3], along with the original and the augmented variant (augVoxNet), calculating both accuracy and mean average precision. The results, together with the first experiment (CGAN), are shown in Table I. The relative accuracy column shows the difference from the original classifier.

The same trend can be seen in these results as in the previous experiment. Both augmented and generative variants outperform the original classifier, whereas the generative network achieves maximum accuracy. Again, the augmented version shows some improvement, however not as significant

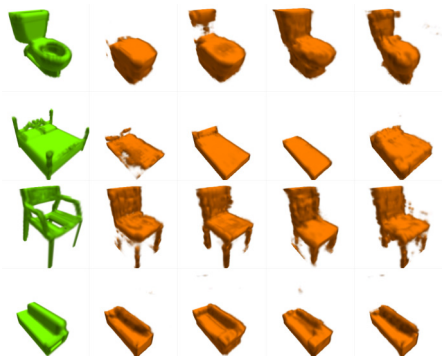


Fig. 1: Output of the conditional generator. The class condition vector has been constructed so that the rows contain examples from these classes (top to bottom): toilet, bed, chair, sofa. The first column contains a random example from the training set.

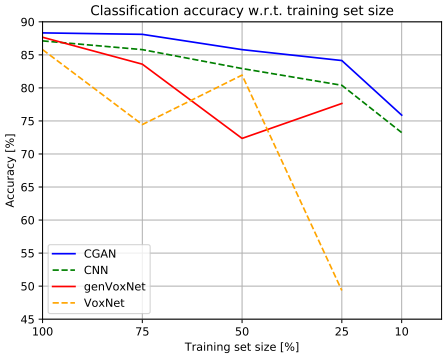


Fig. 2: Classification accuracy of the compared classifiers trained on various sizes of the original ModelNet10 training set.

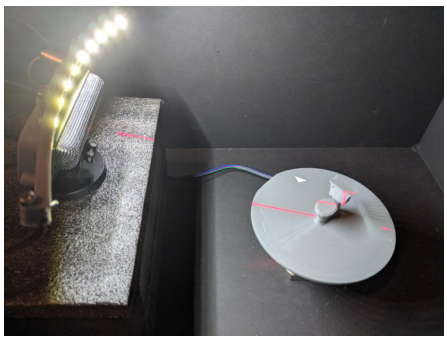


Fig. 3: Photograph of the constructed rotary 3D scanner. The left side shows a 2D camera (PS3 Eye) with a mounted line laser and an LED strip. The other side contains the rotary table driven by a stepper motor.

as the generative classifier. This supports the conclusion of the first (CGAN) experiment. While the improvement in accuracy is slightly lower for VoxNet (2.19% vs 7.43% in the first experiment), it still shows viability and provides benefit for this network.

B. Training set size impact on accuracy

Further evaluation of the original and modified classifiers has been performed to measure the effect of smaller training sets on their performance. We expected that the generative enhancement would improve the classifier's generalization ability, which could increase its performance when trained on less data. Therefore, we designed an experiment, where the classifiers are trained on several training subsets of varying sizes and then validated on the complete testing set. The results are shown as plots in Fig. 2.

In almost all instances, these results show, that the generative enhancement increases the classification accuracy and also slightly diminishes the decline of accuracy when shrinking the set of training data. One anomaly is observed with VoxNet and genVoxNet classifiers trained on 50% subset of the training set. This can likely be attributed to that particular selection of training examples which adversely affects the performance of the generative modification while boosting the original classifier. Nevertheless, it shows that in some rare situations, the generative enhancement can be to the detriment of the classification performance.

C. Rotary 3D scanner – real-world evaluation

To evaluate the viability of the methodology under physical conditions, we designed and built a rotary 3D scanner to scan physical objects and create a real-world validation dataset. This scanner consists of a rotary table driven by a stepper motor, a 2D camera with an LED light and a line laser attached, and an Arduino microcontroller, serving as a communication interface with a PC. A photograph of the assembled device is shown in Fig. 3. The 3D scanning is based on the laser triangulation technique, gathering multiple 3D profiles of the scanned object and collecting them into a single point cloud. Cropped and filtered point cloud is converted into a voxel grid.

10 models from the ModelNet dataset have been 3D printed and scanned with the constructed 3D scanner. The resulting set of voxel models was then classified by CNN and VoxNet classifiers along with their augmented and generative variants. Original and augmented classifiers misclassified 1 out of 10 examples, while both generative variants achieved perfect accuracy. Due to the limited number of samples, it is not possible to draw a general conclusion, however, as the observed trend is in line with previous experiments, it looks like the mechanism of knowledge transfer between the generator and the classifier manifests itself even under real-world conditions.

IV. Conclusion and Future Work

In the presented research, we produced further results supporting the potential of the proposed methodology for generative enhancement of 3D image classifiers. Future research could continue on both theoretical and practical levels. In theory, it would be beneficial to investigate the concrete mechanism of the generative feature knowledge transfer from the generator to the classifier and how those features aid the classifier during validation. On a more practical level, other applications of the proposed methodology could be investigated, considering more complicated classifier architectures.

Acknowledgment

This work has been supported by grant KEGA Implementation of research results in the area of modeling and simulation of cyber-physical systems into the teaching process - development of modern university textbooks – 072TUKE-4/2018.

References

[1] M. Varga, J. Jadlovský, and S. Jadlovská, "Generative enhancement of 3D image classifiers," *Applied Sciences*, vol. 10, no. 21, p. 7433, Oct. 2020.

[2] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference On*. IEEE, 2015, pp. 922–928.

[3] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1912–1920.

Domain wall dynamics in rapid changing magnetic field

[1]Jana HORNIAKOVÁ (2nd year)
Supervisor: [2]Jozef ONUFER

[1,2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]jana.horniakova@tuke.sk, [2]jozef.onufer@tuke.sk

***Abstract*—A new experimental method for the study of single domain wall dynamics in bistable microwires is presented. It enables new possibilities for experimenting with a single magnetic domain wall moving in an inhomogeneous magnetic field. The first experimental results indicate changes in basic dynamic parameters of the domain wall passing through the region of rapid change of magnetic field.**

***Keywords*— amorphous microwire, domain wall, inhomogeneous field**

I. Introduction

Amorphous ferromagnetic glass-coated microwires with positive magnetostriction exhibit specific properties that attract attention from theoretical and application points of view [1-4]. They are prepared by Taylor-Ulitovski method [4-6], which provides a microwire in as quenched state with characteristic mechanical stress distribution [7]. This stress distribution combined with the magnitude and sign of magnetostriction cause a typical domain structure to create. The domain structure of an amorphous glass-coated microwire with positive magnetostriction consists of an axially-magnetized core with radial domains in the outer shell, while most of the volume is occupied by the axially-magnetized core [8]. A closure domain structure is formed at the microwire ends due to a stray field. The hysteresis loop of these microwires is rectangular. That is caused by a characteristic type of magnetization reversal which starts with the release of a single domain wall (DW) from the microwire end, and then propagates along the microwire as a large Barkhausen jump. Using various experiments, the dynamics of a single DW can be studied.

Most of these works deal with DWs, which propagate in constant homogeneous magnetic field. So far, this type of material has not received much attention to the study of DW dynamics in an inhomogeneous magnetic field. One type of experiment which makes it possible to study this problem by trapping and injecting single DWs in magnetic wire by local fields, was presented in [9]. Another experiment [10], is based on trapping DW in a local artificial potential well with controlled shape and subsequently the conditions for releasing the wall from this well were studied.

In this work we present a new experimental method for the study of single DW dynamics in bistable microwires in an inhomogeneous magnetic field. Results of experiment are also discussed.

II. Experimental

Amorphous glass-coated $Fe_{77.5}Si_{15}B_{7.5}$ microwire with positive magnetostriction and unidirectional effect, prepared by means of the Taylor-Ulitovski method, was used in the experiment. The length of the sample used was 12.5 cm with a metallic core radius of 7.5 μm and glass coating thickness of 7.5 μm.

We consider experiment schematically depicted in Fig.1. The bistable microwire is placed in two coaxial magnetizing coils. The first one is a long solenoid (So) and the second one is long and thin solenoid (So1). The region close to the end of the solenoid So1 is a source of inhomogeneous magnetic field. Solenoid So creates the field H and solenoid So1 creates the field H_1. Pick-up coil with regularly wounded turns (PuC) is used to monitor movement of a single DW along the microwire.

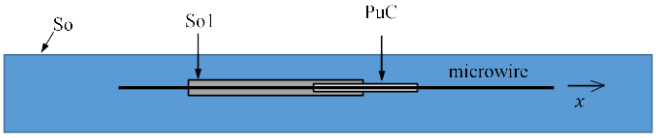


Fig. 1 Experimental set-up

If a single DW propagates along the microwire the equation of motion of this wall is

$$m\frac{dv}{dt} = 2\mu_0 M_s[H + H_1(x) - H_0]A_\perp - \beta v, \qquad (1)$$

where m, v are inertial mass and velocity of the wall, respectively, β is damping coefficient, $A_\perp$ is area of projection of the wall to the microwire cross-sectional plane, M_s is saturation magnetization and H_0 coercive field which we will consider constant and small in further calculations.

For constant homogeneous field H and for $H_1 = 0$ Eq. 1 has a form

$$0 = 2\mu_0 M_s(H - H_0)A_\perp - \beta v \qquad (2)$$

After simple rearrangement

$$v = \frac{2\mu_0 M_s A_\perp}{\beta}(H - H_0) = S(H - H_0) \qquad (3)$$

where $S = \frac{2\mu_0 M_s A_\perp}{\beta}$ is wall mobility. This quantity can be determined from experimental $v(H)$ dependence.

Using wall mobility and after simple rearrangement the Eq.1 can be written in the form

$$\frac{dv^2}{dx} = \frac{4\mu_0 M_s A_\perp}{m}\left[H + H_1(x) - H_0(x) - \frac{v}{S}\right] \quad (4)$$

After substitution $y = v^2$

$$\frac{dy}{dx} = \frac{4\mu_0 M_s A_\perp}{m}\left[H + H_1(x) - H_0(x) - \frac{\sqrt{y}}{S}\right] \quad (5)$$

If we put $x = 0$ at the right end of solenoid So1 the field produced by this solenoid can be expressed as

$$H_1 = I\frac{z}{4l}\left[\frac{x}{\sqrt{R^2 + x^2}} + \frac{2l - x}{\sqrt{R^2 + (2l - x)^2}}\right] \quad (6)$$

Typically experimental $v(H)$ dependences are not linear (Fig.2) as it is predicted by Eq. 3 if wall mobility is considered a constant parameter. Moreover in many cases the value of the field H_0 derived by extrapolation of experimental $v(H)$ dependences is negative. All these facts indicate that wall mobility has to be considered as dependent on applied field (or wall velocity). The reason can be changes in wall dimension/shape in different applied fields [11]. There are two parameters in Eq. 5 which can change as the wall changes its dimension/shape. The first one, already mentioned, is wall mobility S and the second one is wall inertial mass m. Wall mobility versus applied field ($S(H)$) dependence can be determined from experimental $v(H)$ dependence using Eq. 3.:

$$S(H) = \frac{v}{H - H_0} \quad (7)$$

Value of the field H_0 in Eq. 7 can be guessed from low field measurement of $v(H)$ dependences and in many cases it can be neglected for high field regions. Wall mobility expressed by Eq. 7 can be also expressed as a function of wall velocity and the function $\frac{1}{S}(v) = \frac{1}{S}(y)$ can be substituted in the right side of Eq. 5. To predict changes in DW mass is very complex task. Probably the simplest approach can be based on the following idea. Shortening of the wall results in the decrease of both wall mobility and wall mass. So as a first approximation the relation $\frac{1}{m}(v) \sim \frac{1}{S}(v)$ can be considered.

Typical $v(H)$ dependences for a sample with unidirectional effect are shown in Fig. 2.

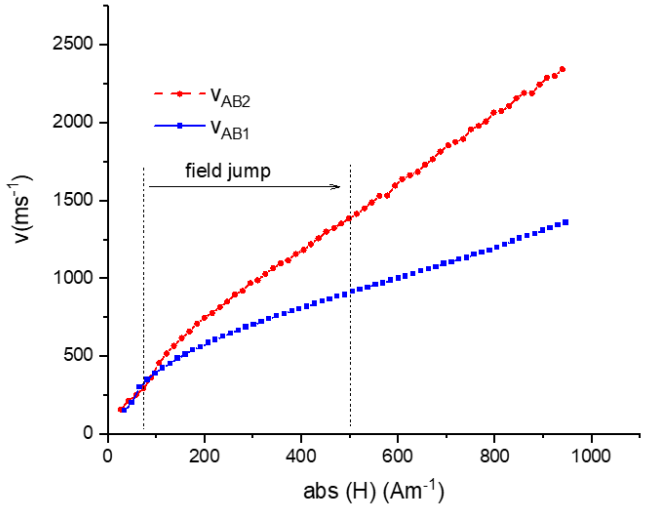


Fig. 2 Velocity versus magnitude of the applied field for sample with unidirectional effect (fast wall v_{AB2}, slow wall v_{AB1}). DW propagates along the same path and in the same direction (where indexes A, B correspond to the ends of the sample) but it is driven by opposite field.

As can be seen in Fig. 3, the jump in velocity appears when the DW propagates through the region of rapid change of the magnetic field. If the DW propagates in homogeneous field, its velocity is approximately constant. When the DW propagates through the region of inhomogeneous magnetic field (jump), it is forced to change its parameters in very short time. Our hypothesis is, that DW can not accommodate to such rapid change immediately. So, gradual adjusting of parameters of the DW is assumed. From further analysis of measured data we should be able to provide information about changes in the DW mobility or DW inertial mass and consequently deduce properties of the DW geometry and inertial properties (relaxation time, distance etc.).

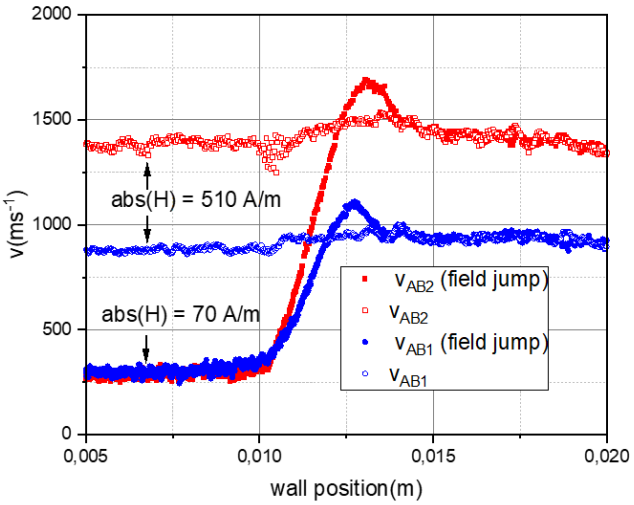


Fig. 3 Position dependencies of DW velocities for DW propagating in homogeneous field (empty symbols) and for DW, that propagates in inhomogeneous field jump from 70A/m to 510 A/m (full symbols).

REFERENCES

[1] M. Vazquez, H. Kronmuller, S. Parkin (Eds.), Handbook of Magnetism and Advanced Magnetic Materials, vol. 4, John Wiley and Sons, Chichester, UK, 2007, pp.1–34.

[2] A. Zhukov, V. Zhukova, Magnetic Properties and Applications of Ferromagnetic microwires with Amorphous and Nanocrystalline Structure (Nova), Science Publishers Inc, Hauppauge, NY, 2009, p.162.

[3] M. Vázquez, H.Chiriac, A. Zhukov, L. Panina, T. Uchiyama, On the state of the art in magnetic microwires and expected trends for scientific and technological studies, Phys. Status Solidi A 208 (2011), pp. 493–501.

[4] M. Kladivová, J. Ziman, Properties of a domain wall in a bi-stable magnetic microwire, J. Magn. Magn. Mater. 480 (2019), pp. 193-198.

[5] H.Chiriac, T. A. Ovari, Amorphous glass covered magnetic wires: preparation, properties, applications, Prog. Mater. Sci. 40 (1996), pp. 333–407.

[6] S. V. Larin, A. Torcunov, A. Zhukov, J. González, M. Vazquez, L. Panina, Preparation and properties of glass-coated microwires, J. Magn. Magn. Mater. 249 (2002), pp. 39–45.

[7] H. Chiriac, T.A. Óvari, G.h. Pop, Internal stress distribution in glass-covered amorphous magnetic wires, Phys. Rev. B 52 (1995) 10104.

[8] R.Sabol, M. Rovnak, P.Klein, Mechanical stress dependence of the switching field in amorphous microwires, IEEE Transactions on Magnetics 51 (2015) 1-4.

[9] M. Vázquez, G. A. Basheed, G. Infante, R. P. Del Real, Trapping and injecting single domain walls in magnetic wire by local fields, Phys. Rev. Lett. 108 (2012) 037201.

[10] J. Ziman, V. Šuhajová, M. Kladivová, Magnetic domain wall dynamics in an inhomogeneous magnetic field, Physica B: Condensed Matter. 407 (2012) 3905-3909.

[11] J. Horniaková, J. Onufer, J. Ziman, P. Duranka, S. Samuhel, Changes in geometry of propagating domain wall in magnetic glass-coated bistable microwire, accepted for publication in J. Magn. Magn. Mater.

Thermal Properties of Substrates for Face-down Technology

[1]*Tomáš LENGER (4th year)*
Supervisor: [2]Alena PIETRIKOVÁ

[1,2]Deptartment of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]tomas.lenger@tuke.sk, [2]alena.pietrikova@tuke.sk

***Abstract*—This paper is a summarization of the last year of post gradual study. This paper is focused on lamination of multilayer PCBs. Substrate material with low T_g as well as substrate material with high T_g were used. Coefficient of thermal expansion is one of the key parameters and it is very important for the stability of substrates for embedding components. Thermo-mechanical analyses was used for determination of the coefficient of thermal expansion of used substrates This paper analyses thermal properties of low T_g and high T_g substrate materials, which can be used for creating multilayer PCBs with embedded components realized by face-down technology.**

***Keywords*—embedded components, prepreg, multilayer PCBs,**

I. Introduction

Development of electronics is putting pressure on designers to place more electronic components on the small Printed Circuit Boards (PCBs). Miniaturization of PCBs based on the reducing trace width, isolation gaps, or the dimensions of components packages is reaching its physical limits. Embedding components into substrate opens the door to new opportunities in miniaturization of electronic devices. Embedded components may slightly increase the thickness of PCB but allow marked reduction of PCB area. Furthermore, embedded components can improve electrical properties or reliability of PCBs [1, 2, 3].

Electronic components can be embedded by many technological processes. Embedded components can be created as planar resistor or capacitors, electronic components in Surface Mount Device (SMD) package or even unpackaged chips. There are a few different technologies for embedding components which are using different technological steps [1, 3].

Face-down technology represents a new technology that has the advantage of miniaturizing of the PCBs dimensions and improving the electrical performances. Principle of application this technology lies in thermal curing procedure of B-stage reinforced dielectric (multi-ply prepreg) in combination with pressure. Different thicknesses, depending on the assumed component thickness, can be achieved by two-stage laminating several thin layers of prepreg forming the core laminate. However, these glass fibre-reinforced composites are not resistant to failure risks due to thermo-mechanical loading during manufacturing and operation in electronics [4, 5, 6]. Before implementing this technology into the mass production, it is needed to analyze thermo-mechanical properties of this multiply nonstandard laminated epoxy based sandwich structure.

II. Lamination of Multilayer PCBs

This research was focused on thermal properties of substrate material, which are used for lamination of multilayer PCBs. Laminate materials with low T_g and high T_g as well as prepreg materials with low T_g and high T_g were used for this research. Basic properties of this materials are in Table 1.

TABLE 1
Basic Properties of Used Materials [5, 6]

Type	Glass cloth style	Catalogue thickness of 1 layer (mm)	T_g (°C)	T_d (°C)
DE 104 (laminate)	2116	0.565	135	315
PCL370HR (laminate)	7626	0.565	180	340
DE 104 (prepreg)	2116	0.105	135	315
DE 104 (prepreg)	7626	0.197	135	315
PCL370HR (prepreg)	7626	0.172	180	340

First step of the two step lamination creates solid structure of prepreg layers which can be manipulated or drilled. This structure allows realization of vias or holes for embedding of components. This approach has big potential for application in components embedding by face-down technology.

Number of layers was chosen to achieve approximately the same thickness of embedded samples. Every prepreg material was laminated by one step lamination (samples marked as ”A”). Laminate materials and prepreg materials were also laminated to sandwich structures (samples marked as ”M”) which were laminated by two-step lamination (Table 2).

TABLE 2
Types of Samples Used in Experiment

Sample	Material	Layers	T_g (°C)	Thickness (mm)
L1	DE104 (laminate)	1	135	0.56
L2	PCL370HR (laminate)	1	180	0.56
1A	DE104 (prepreg 2116)	10	135	1.21
2A	DE104 (prepreg 7626)	6	135	1.27
3A	PCL370HR (prepreg 7626)	6	180	1.36
M1	Combination of L1 and 1A	10+2	135	2.22
M2	Combination of L2 and 1A	10+2	135-180	2.44
M3	Combination of L2 and 3A	6+2	180	2.45

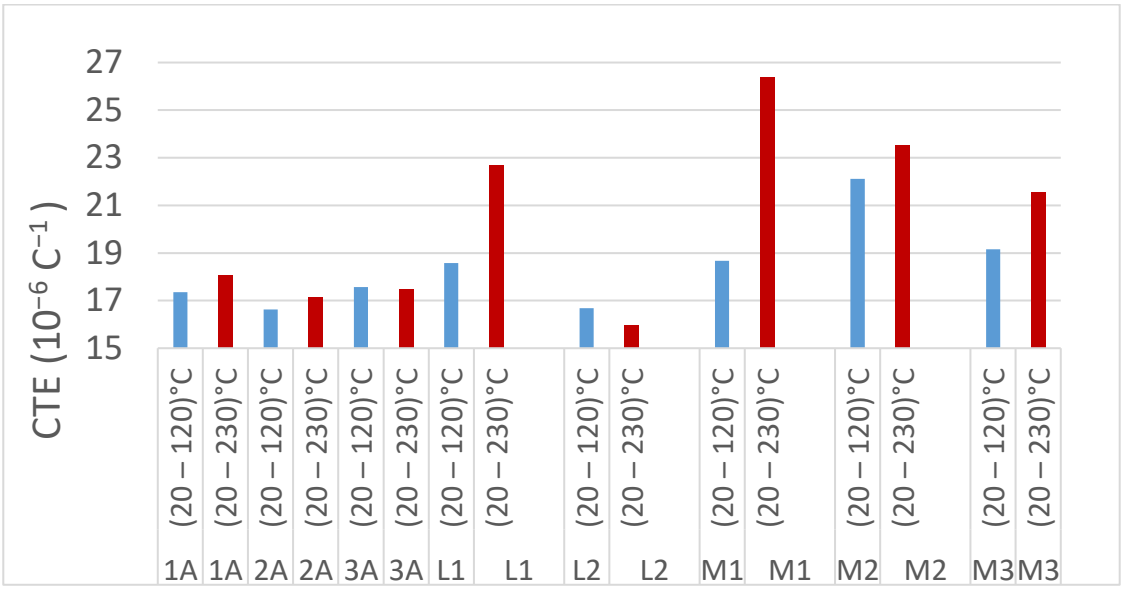


Fig. 1 Coefficient of thermal expansion in "Y" direction below and above Tg.

III. Thermal properties of Multilayer PCBs

TMA was used to characterize linear expansion and softening points of materials by applying a constant force to a specimen while varying temperature. From Figure 1 it is evident that there are no big differences in the CTE (Coefficient of Thermal Expansion) in "Y" direction between multi-ply prepreg laminates. Laminate L1 and sandwich structure based on low Tg (M1) are characterized by the evident difference below and above Tg. The CTE does not show a significant difference in the case multilayer prepreg laminates whether fine fibers cloth or thicker fiber cloth (1A and 2A).

When using higher Tg based materials, there is no significant difference between prepreg laminate and L2 laminate. In the case of the sandwich structure M3, the CTE increased slightly but there was no significant difference in the CTE value in the area below and above the Tg.

Irreversible dimensional change of solid samples after heating and cooling cycle are in Figure 2.

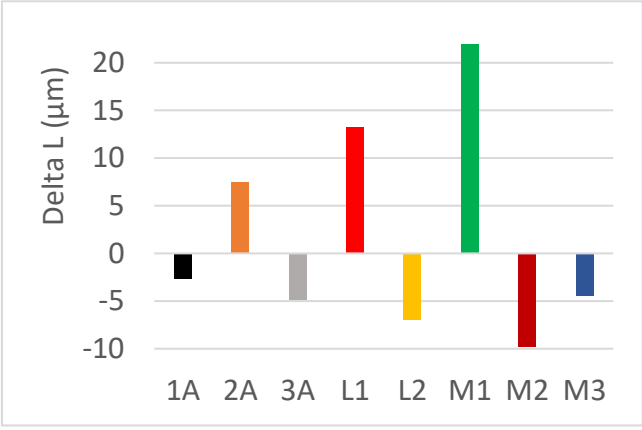


Fig. 2 Comparison of irreversible dimensional change of sample after heating and cooling cycle.

The dilatometric analysis shows that the samples exhibit at least irreversible dimensional changes – expansion (positive change) and shrinkage (negative change) after heat loading at 20 °C. The biggest irreversible change of expansion was observed at a sandwich structure based on low Tg (M1). Sandwich structure based on high Tg (M3) reveals the biggest thermal stability because of minimal of irreversible elongation (shrinkage). Irreversible dimensional changes related to residual stresses and mechanical properties. All these properties are time dependent and they are related to the history of thermal treatment. We noticed visible changes in the sandwich M2 where internal stresses (distortion) occurred at the interface of two types of materials (low Tg and high Tg) during the heat treatment. The result of this is a negative irreversible elongation (shrinkage).

IV. Conclusion

The sandwich structures meet requirements for advanced structural applications in electronics because the possibility to eliminate the distortion, suppression of irreversible dimensional change and warpage effect. Sandwich structure with higher Tg has markedly higher dimensions' stability than sandwich structure with lower T_g. The higher T_g means better thermal and dimensional stability of the substrate.

References

[1] Hyunho, K. "Device Embedded Substrate with Embedding Multi Passive Components for Camera Module." In Pan Pacific Symposium 2015 Proceedings, Kauai, Hawaii, 2015. pp. 242-247. ISBN: 978-1-63439-931-9.

[2] Boettcher, L. et. Al. "Power Electronics Packages with Embedded Components - Recent Trends And Developments." In SMTA International 2013 Proceedings, Fort Worth, Texas, USA, 2013. s. 337-342. ISBN: 978-1-62993-291-0.

[3] Manesiss, D. et al. "Embedding technologies for heterogeneous integration of components in PCBs-an innovative modularization approach with environmental impact" In 21st European Microelectronics and Packaging Conference and Exhibition, Warsaw, Poland, 2017.

[4] Zweben, C., W. S. Smith, and M. W. Wardle (1979), "Test methods for fibre tensile strength, composite flexural modulus, and properties of fabric-reinforced laminates", Composite Materials: Testing and Design (Fifth Conference), ASTM International.

[5] Isola Group, Isola DE104 Datasheet

[6] Isola Group, Isola PCL370HR Datasheet

The GUI design in MATLAB environment for optimization parameters of spiral planar coil

[1]*Peter HRABOVSKÝ (3rd year)*
Supervisor: [2]Ján MOLNÁR

[1,2]Department of Theoretical and Industrial Electrical Engineering, FEI TU of Košice, Slovak Republic

[1]peter.hrabovsky@tuke.sk, [2]jan.molnar@tuke.sk

***Abstract*—This article deals with the design and concept of a system for optimizing the parameters of a spiral planar coil. Next the procedure for obtaining the parameters of the spiral planar coil using simulations in the COMSOL Multiphysics program is described. A database created in Microsoft Excel software was used to store the parameters obtained from the simulation. The process of processing data from a database using by MATLAB was also described. The last part describes the GUI design created in the MATLAB environment.**

***Keywords*—COMSOL Multiphysics, database, Excel, GUI design, MATLAB, spiral planar coil.**

I. Introduction

Today, we encounter a large number of electrical equipment every day. They help us to fulfill our work tasks, meet everyday needs and make our lives easier in different ways. One of these devices is a 3D printer. In recent decades, 3D printing has evolved from a sophisticated process that requires expensive machines (used only by larger companies) to a significant number of relatively inexpensive open-source projects, allowing a much wider audience to use this technology. Today, an ordinary person can buy a 3D printer, create an object in CAD software, or download it from an Internet database, and materialize the model in the comfort of their own home with their own and inexpensive 3D printer. With the advent of new print materials, opportunities to print unprecedented objects increase. Developments in the field of electrically conductive printing materials with admixtures of graphene, carbon or copper open up a wide range of applications in industrial electrical engineering. It allows the printing of sensors and simpler circuits, it can change the entire standard of electronic devices that are now capable of working in three-dimensional space instead of conventional PCB [1].

Given that the coil is an essential part of almost every electronic device, we focused on the production of the coil using 3D printing using conductive material. Of course, the coil can usually be designed in a CAD program and then printed on a suitable 3D printer. However, if we need to create multiple coils with different parameters and shapes, it is advisable to have a system that will make easier this process. On the basis of our proposed system, the user will be able to design a specific spiral planar coil for a given application in a simpler and faster way. We are currently working on creating a system that can create a 2D model based on the parameters obtained from a large number of simulations and then a 3D model of a planar coil of a given shape with exact dimensions. Such a model could then be printed directly using a 3D printer. The following chapters will briefly describe the system concept and GUI design for this system in the MATLAB environment.

II. The System Concept

On the Fig. 1 we can see a diagram of a proposed system concept using three software.

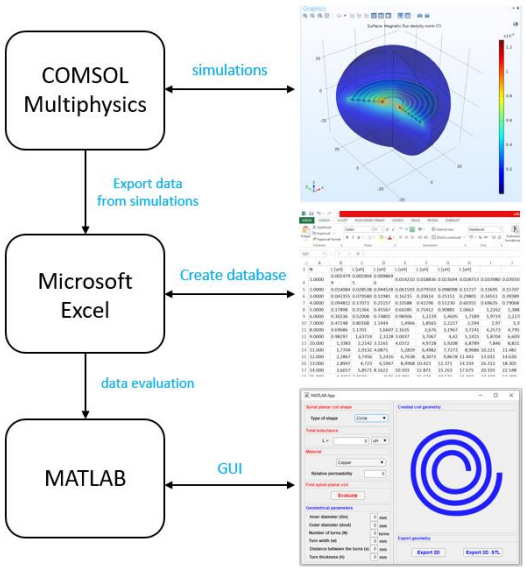


Fig. 1 The diagram of a proposed system concept

There are several procedures for calculating the coil inductance. In the literature, we can find several analytical expressions such as Bryan's formula, Wheeler's formula, Mohan's formula etc [2][3]. However, the analytical approach may be rather complex and it is sometimes difficult to estimate it mainly with increasing complexity of coil geometry. Therefore, the use of the finite element method (FEM) is a suitable alternative for solving these calculations. There are several software that use this method. One of them is a simulation software called COMSOL Multiphysics.

A. COMSOL Multiphysics - simulations

A large number of simulations are created with COMSOL Multiphysics software. With use of to the simulations of various spiral planar coils, we calculate the inductance values and all the necessary parameters for subsequent processing. The Fig. 1 also shows a simulation of a planar spiral coil of circular shape [1].

B. Microsoft Excel - create database

Then the data obtained from the simulations are exported to a software called Microsoft Excel. A database of the spiral planar coils of different shapes with different parameters will be created. This process is shown in the Fig. 1. This database still needs to be processed and appropriately implemented with next software.

C. MATLAB – data evaluation and GUI

The MATLAB program will be used for subsequent processing of data from the MS Excel database. A Graphical user interface (GUI) will also be created in this environment. The next part of this article will describe the GUI design in MATLAB environment [4].

III. Graphical User Interface

Graphical user interface was created using by MATLAB environment using the App Designer function [4]. On the Fig. 2 we can see the GUI of the application.

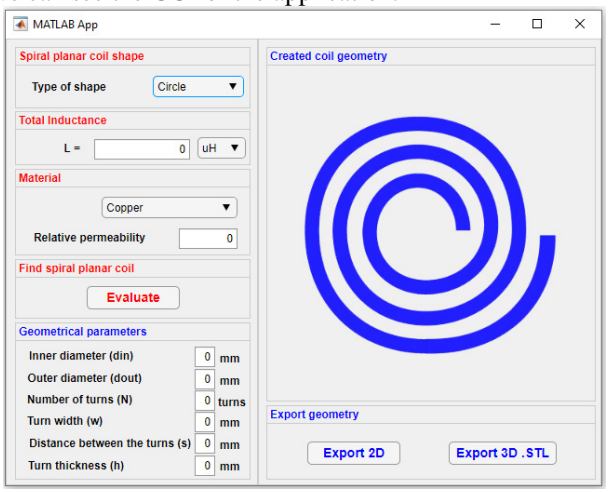


Fig. 2 Graphical user interface created in MATLAB

The GUI is divided into input and output fields. All panels marked in red are used as input parameters. The output parameters are displayed in blue panels. The first option in the application is to select the desired shape of the spiral planar coil named panel *Spiral planar coil shape*. We have a choice of four basic options: circular, square, hexagonal, octagonal.

Then, the user selects the desired inductance of the coil whose parameters he wants to determine. This option is selected in a panel called *Total inductance*.

Next there is a choice of *Material* from which the individual turns of the coil consist. Copper is used by default. The implementation of electrically conductive material intended for 3D printing is also planned.

After entering all the necessary parameters, it is necessary to press the button called *Evaluate*. The program then looks for the most suitable coil depending on the specified parameters.

Then, the values of the spiral planar coil are displayed in the output part called *Geometrical parameters* based on the entered input parameters. These parameters are: Inner diameter (d_{in}), Outer diameter (d_{out}), distance between the turns (s), turn width (w), turn thickness (h), number of coil turns (N). The individual parameters are shown in the Fig. 3. The parameters are shown for four shapes of planar spiral coils square, hexagonal, octagonal, circular.

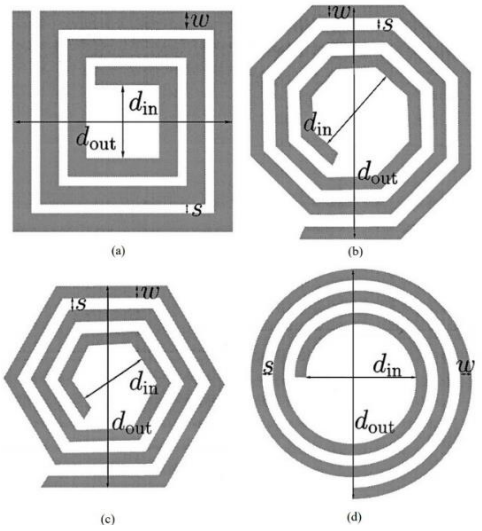


Fig. 3 Type of inductor realization: (a) square, (b) hexagonal, (c) octagonal, (d) circular [2]

In the right part of the program called *Created coil geometry*, a coil with the desired shape and parameters is drawn. then this coil can be exported in the *Export geometry* section, as a 2D model or as a 3D model in .STL format, which will serve as a template for 3D printing of a real spiral planar coil.

IV. Conclusion

The aim of this article was to bring to the user an application that can select the optimal parameters of the coil with respect to the required inductance. The user determines what shape of the spiral planar coil he requires (circular, square triangular, hexagonal), and which inductance the coil wants to achieve, he also chooses the maximum and minimum dimensions of the coil. After entering the parameters, the program searches the database for the most optimal coil for the specified inductance and lists the parameters of the coil and also draws its shape. Then the program should be able to export a 2D model of the drawn coil and also a 3D model of the .STL format. The 3D model created in this way could greatly facilitate the production of a spiral planar coil with the help of a 3D printer.

The whole concept described in this article can be improved by using COMSOL packages from the LiveLink package, namely LiveLink™ for Excel® and LiveLink™ for MATLAB®. However to use these features an extended version of the license is required for all programs mentioned.

References

[1] P. Hrabovský. “Spiral technical coil simulation and parameter verification” In: SCYR 2020 : Nonconference Proceedings of Young Researchers. - Košice (Slovensko): Technická univerzita v Košiciach. pp. 176-177, 2020.

[2] Jonsenser. Zhao, "A new calculation for designing multilayer planar spiral inductors", EDN (Electrical Design News), vol. 5.14, pp. 37, 2010.

[3] S.S. Mohan, M. del Mar Hershenson, S. P. Boyd and T. H. Lee, "Simple accurate expressions for planar spiral inductances," in IEEE Journal of Solid-State Circuits, vol. 34, no. 10, pp. 1419-1424, Oct. 1999.

[4] MATLAB® - MATLAB® App Building R2020b. [online]. Available online at <https://www.mathworks.com/help/pdf_doc/matlab/creating_guis.pdf>

Importance of Normalization in Meta-Learning

[1]*Andrinandrasana David Rasamoelina (2nd year),*
Supervisor: [2]*Peter Sinčak*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]andrijdavid@tuke.sk, [2]peter.sincak@tuke.sk

***Abstract*—Normalization techniques play a dominant role in the training of deep neural networks. It is one of the reasons that helps deep neural networks achieve better results in those recent years. It enables the use of higher learning rates and accelerates the learning process itself. Due to this success, multiple normalization techniques were developed such as Batch Normalization, layer normalization, Weight normalization, Spectral normalization, Instance normalization. However, in a meta-learning setting, some of those normalization techniques can be ineffective or go against the meta-learning principle itself. In this paper, we explore the effect of normalization techniques on different meta-learning algorithms and review the most commonly used normalization techniques.**

***Keywords*—Meta-Learning, Few-shot learning, Normalization**

I. Introduction

Deep learning models have achieved state-of-the-art performance on visual recognition tasks such as image classification. This strong performance, however, relies on abundant labeled instances with diverse visual variations. The cost of human annotation as well as the scarcity of the data in certain applications limit the usability of such a concept. In contrast when humans learn new skills they rarely start from scratch. Usually, they start from skills learned from their past experience in related tasks, reuse approaches that worked well before, and focus on what is likely worth trying based on that experience [1]. Consequently with every new skill learned, learning new skills becomes easier, as it requires fewer examples and less trial-and-error. We, therefore, leverage our past experience to learn new expertise; We learn to learn. The field of meta-learning tries to apply those concepts to Deep Neural Networks.

Meta-learning algorithms, therefore, aim to learn to perform a range of tasks, by training on a limited number of tasks. This allows meta-learning algorithms to learn to learn on its own. That family of algorithms is commonly used to solve Few-shot learning problems, where learning to generalize to unseen classes during training is of prior importance. This is because of the natural ability of meta-learning algorithms to generalize the representation of tasks using limited input information. Consequently, a good meta-learner is capable of adapting itself to situations previously not encountered during training. Eventually, the model should be able to accomplish new tasks. Hence the term, learning to learn. These intended tasks can range from a wide class of problems which can include supervised, unsupervised, or even reinforcement learning [2–7]. A common point to all kinds of neural networks is the need for normalization. Normalizing is often applied to the input, but it also takes place within the network [8]. In this article, we focus on normalization within the network. Those normalization techniques are important as it enables the training of neural networks at a higher speed and achieve higher result [9]. Unfortunately, normalization techniques haven't been explored in meta-learning as much as in deep learning.

II. Related Work

Problems of exploding or vanishing gradient in gradient-based learning are well-known and have normalization has been shown to provide an efficient way to alleviate such problems. Understanding whether batch normalization is the optimal technique is a current challenge not only for the scientific community in meta learning, but also for the wider scientific community in low intelligence [10]. Batch Normalization provides a method of normalization at the batch level that centers and scales through the entire mini-batch by mean and variance and then rescales the result [9]. Normalization efficiency can be reduced if the batch size is not appropriate to support the evaluation of statistics. Batch Renormalization is introduced in [11], where the proposed solution to solve the problem of degradation is to adding more learnable parameters to BN.

Bronskill et al. [12] stated in their research that the hierarchical nature of the meta-learning system raises many challenges that can make traditional batch normalization ineffective. In their research, they address the problem of batch normalization, include the comparison of alternative approaches for meta-learning scenarios, and present a novel approach called TASKNORM, which outperforms all common approaches on 14 datasets.

III. Why do we need normalization?

Normalization provides multiple benefits to neural network. In practice, a normalization mechanism aims to normalize the output of a given layer such that the vanishing gradient problem can be suppressed and hence to reduce the oscillation in the output distribution. It reduces the Internal Coviarance Shift of the neural network [9]. It reduces the Lipschitzness [13] of the loss and the gradients. It participates to the regularization of the model.

IV. Current Normalization Techniques

Batch Normalization (BN) is a technique that normalizes the mean and variance of each of the features at every level of representation during training. The technique involves normalization of the features across the examples in each mini-batch. Empirically it stabilize the gradient (less exploding

or vanishing values) and batch-normalized models appear to overfit less.

Instance Normalization was first introduced by Ulyanov et al. [14]. It was specifically tuned for image style transfer. It is also broadly used in Generative Adversarial Network architectures.

Layer normalization was introduced by Ba et al. [15]. It computes the mean and variance used for normalization from all of the summed inputs to the neurons in a layer on a single training case. Unlike batch normalization, layer normalization performs exactly the same computation at training and test times. The key feature of layer normalization is that it normalizes the inputs across the features. In contrast, in layer normalization, the statistics are computed across each feature and are independent of other examples.

Batch renormalization was proposed by Ioffe [11]. Batch renormalization use a moving average of the mini batch statistics. This is because a moving average provides a better estimate of the true mean and variance compared to individual mini-batches. A downside of this method is that the ability to back-propagate through those statistics is not tractable.

Unfortunately, batch renormalization's performance still degrades when the batch size decreases (though not as badly as batch normalization), meaning group normalization still has a slight advantage in the small batch size regime.

Batch Instance Normalization is an extension of Instance Normalization. Introduced by Nam and Kim [16], BIN learns to selectively normalize only disturbing styles while preserving useful styles. Batch-instance normalization attempts to deal with this by learning how much style information should be used for each task and feature map (channel).

In other words, batch-instance normalization is just an interpolation between batch normalization and instance normalization.

Weight normalization reparameterizes the weights $\boldsymbol{w}$ of any layer in the neural network. Introduced by Salimans and Kingma [17], Weight normalization separates the norm of the weight vector from its direction without reducing expressiveness. This has a similar effect to dividing the inputs by the standard deviation in batch normalization.

As for the mean, the authors of the paper proposed using a method called "mean-only batch normalization" together with weight normalization. Basically, they subtract out the mean of the minibatch but do not divide by the standard deviation. Compared to the standard deviation, the mean apparently has a "gentler" noise thanks to the law of large numbers, so weight normalization can still work in settings with a smaller minibatch size.

Switchable Normalization [18] was the first meta-normalization techniques. It combines Batch Normalization, Instance Normalization, Group Normalization and Layer Normalization for each individual layer.

Transductive Normalization

Normalization techniques like Batch Normalization, Layer Normalization, and Instance Normalization assumes that the data distributions from the training set will provide appropriate normalization statistics for the test data. However in meta-learning, our data are only assumed to satisfy that requirement within a specific task. Therefore, in Transductive Normalization running statistical moment are not used for normalization at meta-test time, but instead replaced with the support set statistics. These statistics are used for normalization, both at meta-train and meta-test time. This approach can be used with any previously described normalization techniques and qualified as transductive.

V. Discussion

The neural network used in our experiment is a 4 layers convolution network as in [19]. The table I shows the resulting accuracy using different normalization techniques. We see that Switchable Normalization outperform the other techniques by a considerable margin. This is explained by the fact that the network found a specific set of values to weight the individual normalization techniques. The table II and table III shows the weight discovered by the network at different depth of the network for respectively the running mean and the running variance. We see that those values slightly differ at every layer. We can argue that those weight are tuned at every task and make this approach ***semi-transductive***. Semi-transductive because the running mean, and variance are still computed but they are weighted properly at every task level.

Method	**1-shot 5-way Accuracy**
Batch Instance Normalization	0.416000
Batch Normalization	0.430000
Batch Renormalization	0.444000
Instance Normalization	0.460000
Transductive Layer Normalization	0.454000
Switchable Normalization	**0.486000**
Task Normalization	0.440000

TABLE I
1-shot accuracy on MiniImagenet after 80k episodes.

VI. Conclusion and Outlook

In this work we took a deeper look at normalization technique for meta-learning. We tested those various normalization under a meta-learning problem. We discovered that the traditional knowledge of transductive normalization does not hold under certain condition. We therefore discover that a semi-transductive approach can work in a meta-learning setting. In further work, we will propose a generalization of normalization techniques for meta-learning.

Acknowledgment

This research is supported by AI4EU project from the European Union's Horizon 2020 research and innovation programme under grant agreement 825619 2019-2021, and the

Layer	Instance Normalization	Layer Normalization	Batch Normalization
1	0.31487897	**0.36768788**	0.31743312
2	0.16585885	**0.48280904**	0.35133216
3	0.22115272	**0.46229908**	0.3165482
4	0.07319388	0.459545	**0.46726117**

TABLE II
Mean weight discovered by the neural network

Layer	Instance Normalization	Layer Normalization	Batch Normalization
1	0.3113259	0.25332558	**0.43534854**
2	**0.4869882**	0.26878116	0.24423067
3	0.26575363	**0.3904737**	0.3437727
4	0.30311465	**0.38932243**	0.30756292

TABLE III
Variance weight discovered by the neural network

Maria Currie RISE LIFEBOTS Exchange Grant Agreement ID : 824047 2019-2021, and FEIGRANT – 2021 – Increase server computing power to work with state-of-the-art artificial intelligence algorithms.

REFERENCES

[1] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and brain sciences*, vol. 40, 2017.

[2] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *arXiv preprint arXiv:1806.06877*, 2018.

[3] Y. Bengio, T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, and C. Pal, "A meta-transfer objective for learning to disentangle causal mechanisms," *arXiv preprint arXiv:1901.10912*, 2019.

[4] S. Sohn, H. Woo, J. Choi, and H. Lee, "Meta reinforcement learning with autonomous inference of subtask dependencies," *arXiv preprint arXiv:2001.00248*, 2020.

[5] K. Xu, E. Ratner, A. Dragan, S. Levine, and C. Finn, "Few-shot intent inference via meta-inverse reinforcement learning," 2018.

[6] S. Whiteson, "Varibad: A very good method for bayes-adaptive deep rl via meta- learning," 2020.

[7] B. Piot, M. Geist, and O. Pietquin, "Bridging the gap between imitation learning and inverse reinforcement learning," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 8, pp. 1814–1826, 2016.

[8] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[9] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[10] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger, "Understanding batch normalization," in *Advances in Neural Information Processing Systems*, 2018, pp. 7694–7705.

[11] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," in *Advances in neural information processing systems*, 2017, pp. 1945–1953.

[12] J. Bronskill, J. Gordon, J. Requeima, S. Nowozin, and R. E. Turner, "Tasknorm: Rethinking batch normalization for meta-learning," *arXiv preprint arXiv:2003.03284*, 2020.

[13] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer Science & Business Media, 2013, vol. 87.

[14] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[15] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[16] H. Nam and H.-E. Kim, "Batch-instance normalization for adaptively style-invariant neural networks," *arXiv preprint arXiv:1805.07925*, 2018.

[17] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," *arXiv preprint arXiv:1602.07868*, 2016.

[18] P. Luo, R. Zhang, J. Ren, Z. Peng, and J. Li, "Switchable normalization for learning-to-normalize deep representation," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[19] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Advances Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4077–4087.

Modeling and Implementation of Distributed Control Systems for Large Physical Experiments

[1]*Milan TKÁČIK (2nd year),*
Supervisor: [2]*Ján JADLOVSKÝ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]milan.tkacik@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract*—The aim of this paper is to present a newly developed distributed control system for detectors of the ALICE experiment at CERN. It also describes the necessity of control system modeling, before actual deployment. The paper also proposes model types that can be used to represent individual parts of the system in order to create a complex system model.**

***Keywords*—Detector Control System, Distributed Control System, Finite-State Machines, Petri Nets**

I. Introduction

This paper deals with modeling and implementation of Distributed Control Systems (DiCS), focusing on the development of the Detector Control System (DCS) of the ALICE (A Large Ion Collider Experiment) at CERN. It describes part of the newly developed DCS, that will be deployed at ALICE after the ongoing upgrade is finished.

The paper also presents the need for DCS modeling and proposes solution for developing a complex model of the entire system. The DCS is considered to be a Cyber-Physical System that can be modeled as a hybrid system, and thus several model types can be used to represent its behavior.

II. Previous analysis and achieved results in research field

The ALICE is a complex of 18 detectors at the LHC (Large Hadron Collider) at CERN focused on the study of the quark-gluon plasma formed by ultrarelativistic collisions of heavy ions [1]. The safe and stable operation of the individual detectors, as well as the entire experiment, is ensured by the DCS. The DCS is a massive DiCS, developed in close cooperation with individual detectors, as each detector is unique and has its own requirements for operation. The goal of the DCS is to control and monitor detectors, as well as streamline the operation of the experiment in an effort to reduce detector downtime and maximize the number of readout channels, including processing and archiving all parameters needed for effective physical data analysis and reconstruction of individual collisions.

At the end of 2018, the Run 2 period ended and the LHC together with all experiments were shut down (LS2) in order to modernize and maintain the detectors [2]. Completion of the modernization and the beginning of the Run 3 period is expected at the end of 2021. Significant modernization of hardware and software resources during LS2 is taking place within the ALICE experiment, where the ITS (Inner Tracking System) detector is undergoing the biggest change since it will be replaced with entirely new detector electronics, e.g. new Hybrid Integrated Circuit (HIC) containing new generation of silicon pixel chips. The modernization of individual detectors requires the development of a new DCS, mainly due to the large increase in the amount of processed data.

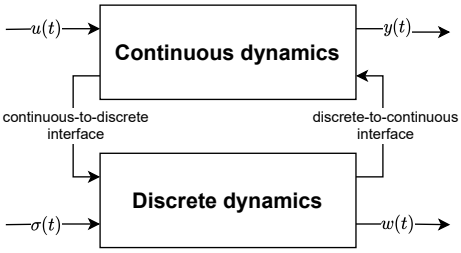


Fig. 1. Interconnection of continuous and discrete dynamics of hybrid systems

Deployment of a new DCS is a complex task that requires finding a suitable distribution of software and hardware resources within the control structure of the system. Due to the size of the DCS, it is necessary to find a balance between the number of servers used and ensuring the necessary performance and stability of the entire system. The necessity of DCS modeling is based on the need to solve this task.

The DCS can be considered a complex Cyber-Physical System that can be modeled as a hybrid system. Hybrid systems are dynamic systems whose behavior exhibits both continuous and discrete dynamics, see Fig. 1. A hybrid system can have one or more continuous dynamics, which can be described by differential equations, where we consider the continuous input $u(t)$ and the continuous output $y(t)$ of the system. In terms of discrete dynamics of hybrid systems, we consider the discrete input $\sigma(t)$ and the discrete output $w(t)$, whereas their relations can be described by discrete models, e.g. Finite-State Machines (FSM) or by Petri Nets (PN) [3].

III. Solved tasks and results

Although the architecture of DCS of individual detectors uses the same concept, their structure is slightly different due to different detector electronics [4]. Fig. 2 shows the DCS architecture of the ITS detector, which can be divided into the central DCS system and four individual subsystems, i.e. frontend system, detector safety system, power system, and cooling system [5].

The development of new detector electronics brings with it the need to ensure compatibility with WinCC OA systems of

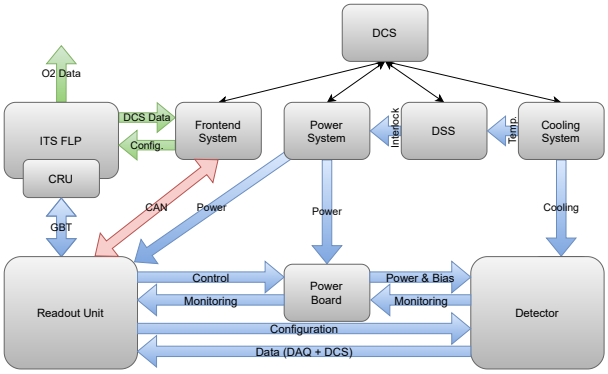


Fig. 2. DCS architecture of the ITS detector

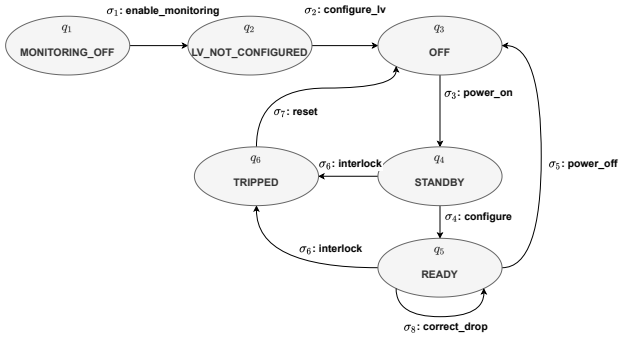


Fig. 3. FSM model for the HIC unit within the central DCS

the central DCS. The communication interfaces and protocols used by the new electronics not only differ from detector to detector, but are also not compatible with the commercial SCADA system WinCC OA. Instead of developing drivers for the WinCC OA, the ALFRED (ALICE Low-Level Front End Device) architecture has been designed, which is implemented as an additional software layer ensuring compatibility between custom electronics and SCADA systems [5].

The implemented FRED (Front End Device) server application is a key part of the ALFRED architecture of the ALICE DCS. It provides translation of raw data obtained from detector electronics into the format of physical quantities, which are necessary to ensure continuous operation of the detector [6]. It also provides the mechanisms needed to implement detector monitoring, control, and protection. At the same time, the FRED is designed so that simpler operations of the detector can be implemented using configuration files without the need for any additional programming. The FRED can control detector electronics using the GBT communication interface, as well as using the CAN bus backup channel [7].

Modeling of the DCS provides the possibility of optimizing the proposed system before its deployment. The created model allows to estimate the duration of individual processes and thus determine the required number of hardware resources needed for each subsystem. The proposed methodology for the DCS modeling uses a combination of PN and FSM discrete models.

PNs are suitable for modeling the lower levels of the DCS, in particular the ALFRED system in terms of modeling a data flow within the individual modules of the system. Using PNs, it is possible to create a model of the ALFRED system at different levels of abstraction, that can be interconnected hierarchically, and thus one block of a higher level of abstraction can represent the entire PN at a lower level of abstraction.

For modeling higher levels of the DCS system, it is more appropriate to use FSM models, as at the level of the central DCS system it is necessary to describe states of the electronics and transitions between them. The employed FSM models are also hierarchical providing several levels of abstraction of the entire system, while the individual levels of FSM models are connected by a strictly defined mechanism, where commands are sent from higher to lower level and states are spread in the opposite direction [8]. Fig. 3 shows an example of the proposed FSM model for the HIC unit of the ITS detector, where states of the model represent the real state of the electronics and inputs of the model represent individual commands from the DCS and events generated by the electronics.

IV. Future Research Steps

Future research steps will involve the continuation of the development of the methodology for modeling large distributed systems using PN and FSM models. The methodology will be used to create a complex model of the DCS, that shall be used to optimize usage of software and hardware resources when deploying the DCS on the production network. Development of the FRED will also continue based on new requests from individual detectors.

V. Conclusion

This paper presented an overview of newly developed DCS at ALICE, as well as proposed an approach for modeling such a complex system. The proposed methodology can also be used for modeling DiCS at DCAI, including mobile robotics applications since the DiCS at DCAI shares the same principles with the DCS at ALICE.

Acknowledgment

This work has been supported by the project ALICE experiment at the CERN LHC: The study of strongly interacting matter under extreme conditions (ALICE KE FEI TU 0222 / 2016) and by grant KEGA Implementation of research results in the area of modeling and simulation of cyber-physical systems into the teaching process - development of modern university textbooks – 072TUKE-4/2018.

References

[1] K. Aamodt, A. A. Quintana, R. Achenbach, S. Acounis, D. Adamová, C. Adler, M. Aggarwal, F. Agnese, G. A. Rinella, Z. Ahammed *et al.*, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 002–S08 002, aug 2008.

[2] M. Bernardini and K. Foraz, "Long shutdown 2@ lhc," *CERN Yellow Reports*, vol. 2, no. 00, p. 290, 2016.

[3] A. J. Van Der Schaft and J. M. Schumacher, *An introduction to hybrid dynamical systems*. Springer London, 2000, vol. 251.

[4] P. Chochula, L. Jirden, A. Augustinus, G. De Cataldo, C. Torcato, P. Rosinsky, L. Wallet, M. Boccioli, and L. Cardoso, "The ALICE detector control system," *IEEE Transactions on Nuclear Science*, vol. 57, no. 2, pp. 472–478, 2010.

[5] J. Jadlovský, A. Jadlovská, S. Jadlovská, M. Oravec, D. Vošček, M. Kopčík, J. Čabala, M. Tkáčik, P. Chochula, and O. Pinazza, "Communication architecture of the Detector Control System for the Inner Tracking System," p. THPHA208, 2018.

[6] P. Chochula, A. Augustinus, P. Bond, A. Kurepin, M. Lechman, J. LÃ, O. Pinazza *et al.*, "Challenges of the alice detector control system for the lhc run3," *ICALEPCS Barcelona, Spain*, 2017.

[7] 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, p. 565, 2020.

[8] G. De Cataldo, A. Augustinus, M. Boccioli, P. Chochula, and L. S. Jirdén, "Finite state machines for integration and control in alice," *ICALEPCS07, Knoxville, USA*, 2007.

Application specific integrated circuits for UWB sensors systems

[1]*Miroslav Sokol (4th year),*
Supervisor: [2]*Pavol Galajda*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]miroslav.sokol@tuke.sk, [2]pavol.galajda@tuke.sk

***Abstract*—This paper introduces designed and realized Application Specific Integrated Circuit (ASIC) for Ultra-Wideband (UWB) sensor systems. The presented ASIC circuits represent front-end circuits and data acquisition and digitalization circuits. These circuits complement the UWB sensor system and make it compatible with many applications. There are mentioned wideband differential amplifiers, wideband directional couplers, wideband witches, and also application-specific high-speed 7-bit AD converter. The analog front end circuit can operate up to 15 GHz and the AD converter which is represented as a mixed-signal circuit is possible to use up to 1.4 GHz sampling frequency.**

***Keywords*—UWB, ADC, ASIC, front-end**

I. Introduction

Nowadays the ultra-wideband (UWB) sensor systems are smaller and cheaper, which meant an expansion of these systems in the 21st century. Electronic miniaturization is the best way to get better, cheaper, and smaller UWB sensors. The UWB sensors system consists of many types of topologies or radiated signals. Commonly the most used types are impulse radars [1], noise radars [2] and FMCW radars [3]. These UWB sensors systems are perspective in many applications and they represent part of the category of sensors capable of non-invasive measurement.

Using non-invasive electromagnetic wave-based measurement it is possible to get information about materials, dimensions, or location of objects or persons. Our research is focused to design and develop UWB sensors based on M-sequence. For the last twenty years, several types of M-sequence UWB sensors have been developed for many applications. It can be operated at few hundreds of MHz to tens of GHz, depends on the application used and semiconductor technology. The M-sequence UWB sensors are used in GPR [1] [4], material analysis [5], [6], person localization and monitoring their vital functions [7], sensors for the medical diagnostic [8] or automotive applications [9].

For each application, the specific hardware is required with a specifically designed front-end ASCI circuit. As part of development, the M-sequence UWB sensor transceiver was designed and implemented, in 0.35 μm semiconductor technology [10]. This technology represents relatively cheap semiconductor technology, in comparison with others [11]. The maximum transit frequency of used technology is f_t= 65 GHz, which allows the design of the ASIC [2] circuit up to 15 GHz. The designed transmitter can operate at 10 GHz baseband frequencies, with 6 GHz usable bandwidth. The implementation of our sensor transmitter to the specific applications requires numerous types of front-end circuits [12]. Also, the developed UWB transceiver is only analog "head", therefore digitalization and signal processing are necessary. For better compatibility and further miniaturization, the design of front-end circuits in the same or related semiconductor technology is recommended. In the future, this will allow the design of the entire system for a specific application on a single chip, system on chip (SoC), or at least in one package as a system in package (SiP). What again will reduce the requirements for PCBs design as well as the cost of the whole system.

II. Front end circuits

The numerous types of UWB differential amplifiers [13], [14], the UWB directional couplers [15], also UWB switches [16] as front-end circuit has been designed. The photography of designed front-end circuits is shown in Fig. 1. The new M-sequence UWB sensor transceiver uses for analog and clock signals differential signaling. In practice, this means a differential signaling of analog signal or digital communication has greater resistance to interference. This also has lower emission to the environment compared to non-differential signaling, which has a positive effect on the elimination of crosstalks. The differential amplifiers were designed for many purposes, for example like receiving amplifier [14], transmitting amplifiers [17] or special purposes like BalUn [3] or UnBal[4].

The differential amplifiers was also designed as part of active wideband directional coupler. Because it is not possible to design a perfect amplifier for all applications several types of amplifiers with various parameters have been designed. It was necessary to take into account the trade-off requirements. For example, is not possible to design an amplifier with high gain and very large bandwidth, one characteristic excludes the other.

The Cherry-Hooper structure was used as main structure for these differential amplifiers. It achieves a good relation between gain and bandwidth in used semiconductor technology. The amplifiers gain ranges from 8 dB to 20 dB and bandwidth ranges from 6 Ghz to 12 GHz (-3 dB). The amplifiers have

[1]Ground penetrating radar
[2]Application Specific Integrated Circuit
[3]BalUn - Balanced to unbalanced
[4]UnBal - Unbalanced to balanced

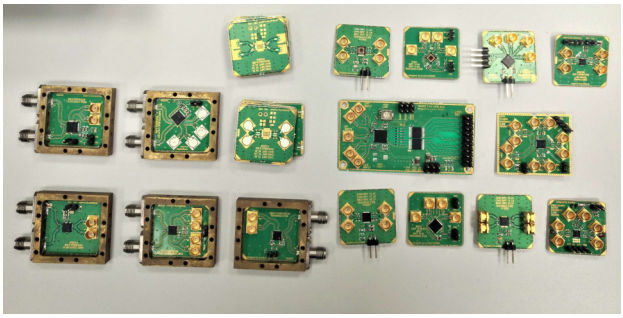


Fig. 1: Several types of designed front-end UWB ASIC circuits assembled on evaluation PCB boards

different types of inputs with impedance from 50 Ω trough 100 Ω differential up to high impedance around kΩ.

For reflectometer measurement several types of active wideband directional couplers have been designed. The couplers are based on the Wheatstone bridge[18]. The Wheatstone bridge can represent three-port or four-port configuration of the directional coupler which belongs to the category of the resistors coupler. The biggest advantage of resistor directional coupler is possibility of integration as ASIC and it is possible to create a very small wideband coupler compared to other couplers. Also it is possible to create active directional coupler, with combination of the Wheatstone bridge directional coupler and differential amplifier. The designed couplers achieve -3 dB to -10 dB forward insertion loss with 0 dB to 5 dB coupling factor and the directivity achieves up to 30 dB. The usable bandwidth of active directional coupler achieves 8 GHz (-3 dB).

The next category of designed front-end circuits is UWB switch . It was primarily designed for sensor calibration unit, but also is possible to use it for switching transmitting and receiving signals. The switch was also implemented, in 0.35 µm semiconductor technology. It is based on CMOS switching principle [19]. The switch has been designed in four port configuration with one common port and three slave ports. The usable bandwidth of the proposed switch is up to 6-GHz (-3 dB).

III. Data acquisition and digitalization circuits

For digitalization and digital data processing of measured UWB sensor data, the analog digital converter (ADC) is necessary. The analog transceiver uses the sub-sampling principle for reducing the sampling speed of the UWB sensor system [2]. Nevertheless, they still operate at hundreds of MHz . Up to now, all designed UWB sensors systems used commercially available AD converters. Besides that, there are more reasons for developing custom made AD converter design. The naked die chips of commercial converters are not available in general for SiP solution and the license of the layout masks of the AD converter from third parties for SoC solution costs hundreds of thousands of euros. On the other hand, the price of custom made application-specific AD converter is significantly lower and you can also make a design in the same semiconductor technology as the analog part of the whole UWB sensor for further SoC solution.

In our case the parallel flash ADC structure has been chosen for its simple working principle [20]. The parallel flash ADC topology is the best option to achieve the highest speed in relatively large semiconductor technology, but its main

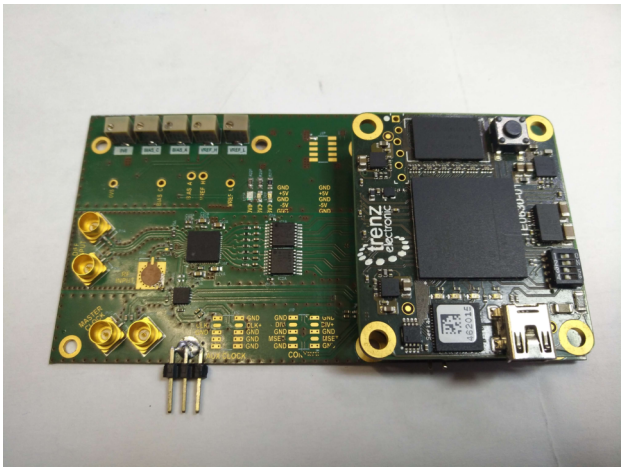


Fig. 2: Designed data acquisition board with custom made ASIC AD converter

disadvantages are large power consumption and large chip area occupation. The design of the presented AD converter has been based on the previous 5-bit [21], and 4-bit AD converter designs [22], [23].

From the requirements of the analog part of the UWB sensor, the 7-bit Flash AD Converter in 0.35 µm semiconductor technology has been designed. Thus in the same semiconductor technology as analog transceiver. The designed AD converter assembled on PCB with FPGA module is shown in Fig.2. The die size is 2x2mm and it was wire-bonded to QFN48 6x6 mm package. At present, its basic parameters from measurements are evaluated. The ENOB [5] of proposed AD converter achieves 6.5-bit at 200 MHz sampling frequency. The ADC input can process analog signal with bandwidth up to 500 MHz. The maximum sampling frequency achieves 1.4 GHz. The maximum INL [6] achieves 2 LSB, and maximum DNL[7] is up to 1.5 LSB. The power consumption of proposed AD converter is 1280 mW at maximum sampling frequency.

IV. Conclusion

The many kinds of ASIC circuits for UWB sensor systems were designed and implemented. For the last few years of development, all necessary front-end circuits have been designed. Namely, there are wideband differential amplifiers, wideband active directional couplers, 4-bit and 7-bit UWB AD converters. Also there are few other circuits that have been designed but they are not presented in this paper, like frequency mixers, VCO oscillators and binary dividers. Future developments will be focused on implementing designed circuits into interesting UWB applications and also developing new ASIC circuits with better parameters.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-18-0373, Scientific Grant Agency (VEGA) under the contract No. 1/0584/20 and and the Slovak Cultural and Educational Grant Agency (KEGA) under the Contract No. 009TUKE-4/2019.

[5] effective number of bits
[6] Integral nonlinearities
[7] Differential nonlinearities

References

[1] V. I. Koshelev, V. P. Belichenko, and Y. I. Buyanov, *Ultrawideband short-pulse radio systems*. Artech House, 2017.

[2] J. Sachs, *Handbook of Ultra-wideband Short-range Sensing*, 2012.

[3] J. Gamec and M. Gamcova, "Experimental Frequency Modulated Continuous Wave radar," in *Radar Symposium (IRS), 2015 16th International*, IEEE, 2015, pp. 1010–1015.

[4] P. Galajda, M. Pecovsky, J. Gazda, and M. Drutarovsky, "Novel M-sequence UWB sensor for ground penetrating radar application," in *2018 IEEE Asia-Pacific Conference on Antennas and Propagation (APCAP)*, IEEE, 2018, pp. 110–111.

[5] M. Kmec, M. Helbig, R. Herrmann, P. Rauschenbach, J. Sachs, and K. Schilling, "Toward integrated µNetwork analyzer," in *Ultra-Wideband, Short-Pulse Electromagnetics 10*, Springer, 2014, pp. 443–451.

[6] M. Žiga, P. Galajda, S. Slovak, and M. Kmec, "Determination of the quality of frying oil based on UWB impedance spectrometer," in *2015 16th International Radar Symposium (IRS)*, 2015, pp. 955–960. DOI: 10.1109/IRS.2015.7226378.

[7] D. Kocur, D. Novák, and M. Švecová, "Multiple Person Localization Based on Their Vital Sign Detection Using UWB Sensor," *Microwave Systems and Applications, SK Goudos, Ed. InTech*, pp. 399–422, 2017.

[8] J. Sachs, M. Helbig, S. Ley, P. Rauschenbach, M. Kmec, and K. Schilling, "Short interfacial antennas for medical microwave imaging," in *2017 International Workshop on Antenna Technology: Small Antennas, Innovative Structures, and Applications (iWAT)*, Mar. 2017, pp. 245–248. DOI: 10.1109/IWAT.2017.7915370.

[9] Z. Qian, T. Wang, and J. Chen, "Compatibility Studies of IMT System and Automotive Radar in the Frequency Range 24.5-25.5 GHz," in *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, IEEE, 2018, pp. 505–508.

[10] P. Galajda, S. Slovak, M. Sokol, M. Pecovsky, and M. Kmec, "Integrated M-Sequence based transceiver for uwb sensor networks," *Radioengineering*, vol. 28, no. 1, pp. 175–182, 2019, ISSN: 1805-9600. DOI: 10.13164/re.2019.0175.

[11] EUROPRACTICE. "Europractice ic service." (2019), [Online]. Available: http://www.europractice-ic.com/.

[12] P. Galajda, M. Pecovsky, M. Sokol, M. Kmec, and D. Kocur, "Recent advances in asic development for enhanced performance m-sequence uwb systems," *Sensors*, vol. 20, no. 17, p. 4812, 2020.

[13] M. Sokol, P. Galajda, S. Slovak, and M. Pecovsky, "Improving parameters of cherry-hooper amplifier using parasitic elements on the layout," in *2019 PhotonIcs & Electromagnetics Research Symposium-Spring (PIERS-Spring)*, IEEE, 2019, pp. 451–456.

[14] M. Pecovsky, M. Sokol, S. Slovak, P. Galajda, and M. Drutarovsky, "Integrated differential amplifier for ultra-wideband electrically short antennas," in *2019 IEEE Radar Conference (RadarConf)*, IEEE, 2019, pp. 1–5.

[15] M. Sokol, P. Galajda, S. Slovak, M. Pecovsky, and M. Drutarovsky, "Design and implementation of microwave circuits in 0.35 μm sige bicmos technology for uwb applications," in *2019 PhotonIcs & Electromagnetics Research Symposium-Spring (PIERS-Spring)*, IEEE, 2019, pp. 48–52.

[16] M. Sokol, *Návrh polovodičového integrovaného spínača pre uwb aplikácie*, 2015.

[17] M. Sokol, P. Galajda, S. Slovak, and M. Pecovsky, "Differential Amplifier Based on Cherry-Hooper Topology for UWB M-sequence Radar," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–5. DOI: 10.1109/RADAR.2019.8835573.

[18] A. Kraus, *Einführung in die Hochfrequenzmeßtechnik: mit 254 Bildern und Tabellen*. Pflaum, 1980.

[19] X. J. Li and Y. P. Zhang, "Flipping the CMOS switch," *IEEE Microwave Magazine*, vol. 11, no. 1, pp. 86–96, 2010.

[20] S. Rapuano, P. Daponte, E. Balestrieri, L. D. Vito, S. J. Tilden, S. Max, and J. Blair, "ADC parameters and characteristics," *IEEE Instrumentation Measurement Magazine*, vol. 8, no. 5, pp. 44–54, 2005, ISSN: 1094-6969. DOI: 10.1109/MIM.2005.1578617.

[21] M. Liptaj, M. Ziga, and P. Galajda, "Design of the 5-bit Flash AD Converter for UWB Applications," in *12th Scientific Conference of Young Researchers*, 2012, ISBN: 978-80-553-0943-9.

[22] M. Sokol, P. Galajda, S. Slovak, and M. Pecovsky, "Design of 4-bit Flash ADC Cell for UWB Sensor Systems," in *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*, 2019, pp. 1–5. DOI: 10.1109/RADIOELEK.2019.8733510.

[23] M. Sokol, P. Galajda, and M. Pecovsky, "Improving input circuits for 7-bit flash ad converter," in *2020 30th International Conference Radioelektronika (RADIOELEKTRONIKA)*, IEEE, 2020, pp. 1–5.

Final Proposal on Modeling and Control for Walking Robots

[1]*Lukáš KOSKA (4^{th} year),*
Supervisor: [2]*Anna JADLOVSKÁ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FFEI TU of Košice, Slovak Republic

[1]lukas.koska@tuke.sk, [2]anna.jadlovska@tuke.sk

***Abstract*—This paper presents the research work and results which were gained over the final year of PhD study in the field of modeling and control of walking robots. Various model configurations have been designed as part of the modeling process, based on the proposed methodology. In the field of control, the process of generation and tracking of trajectories will be presented. The theory of hybrid systems was used to solve research tasks within the experiment ALICE in CERN.**

***Keywords*—hybrid systems, walking models, trajectory generation, detector-control system**

I. Introduction

Human walking stands above all of the other forms of biped locomotion due to the fact that during a major part of the step, the moving body does not occur in the static equilibrium position [1]. The movement of the robot's legs, while passive walking on the inclined plane, is controlled by its own dynamics, powered by the gravitational acceleration [2]. The aim of the control of underactuated walking is to utilize the natural dynamics of the robot, because the control is more efficient and requires less power to the system than the control which tracks the reference trajectory in the fully-actuated system [3].

II. Previous analysis and achieved results

While modeling of walking as such, it is necessary to consider not only the continuous dynamics of the legs but also the discrete events that occur when swing leg impacts the ground. For this reason, it is the most useful to use the concept of the hybrid systems that combine the continuous and discrete dynamics of the model [4].

Methods for modeling and control of walking robots were analysed and appropriate methods were selected.

A crucial feature that arises in the planned motion of the underactuated robotic walking is periodicity, since the steps are periodically repeated [5][6]. Simple benchmark underactuated systems such as cart and pole and acrobot serve to verify control methods for underactuated robotic walking [7]. Several control methods have been applied to be used in control the biped locomotion, most of them use a feedforward control structure, e.g. combining the boundary value problem and optimal control to generate and then track the desired trajectory [5].

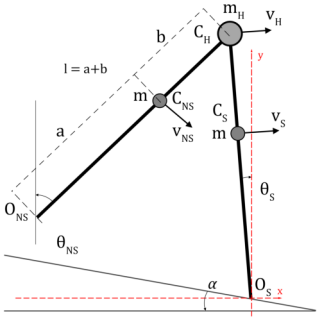


Figure 1. Model of the Compass Gait [8]

III. Solved Tasks

During the final year, methodologies for modeling and control of bipedal walking robots were developed and verified using several models of compass gait-based walking robots.

The developed methodology for modeling walking robots is based on hybrid systems theory. The dynamics of the swing and stance legs is derived using Lagrange equations. The discrete dynamics of the walking robots includes several discrete situations, which can by modeled using the law of conservation of momentum. The methodology was proposed with appropriate software support.

Using the proposed methodology, we generated mathematical models of the compass gait system and its derivatives which include knees, feet and torso. Derived models were implemented in MATLAB.

The Compass Gait Model - (CG), shown in Fig. 1 represents the basic walking model, which is unactuated and walks down a plane inclined by the angle α or model, which is actuated and walks on flat ground. The angle $\theta_{NS}(t)$ in Fig. 1 represents the swing (non-stance) leg and angle $\theta_S(t)$ represents the stance leg. This basic model was extensively studied by Goswami et. al. [9].

The discrete dynamics of the CG model includes only one situation when the swing leg falls on the ground. If the collision condition $s(\boldsymbol{\theta}) : \theta_{NS}(t) + \theta_S(t) + 2\alpha = 0$ is fulfilled, the swing leg becomes stance and stance leg becomes swing leg [10].

Several existing control methods have been used in control the walking robots, most of them use a feedforward control structure. Such a control algorithm for bipedal walking robots can be generated using the partial feedback linearization (PFL) method, since an underactuated system is considered. The

principle of partial feedback linearization is to add such feedback that ensures that the system behaves like a linear system.

The reference trajectories necessary for the control algorithm implementation can be obtained using the model of the passive walk on an inclined plane [8]. Linearizing feedback allows to specify the control law in the form:

$$v(t) = \ddot{\theta}^*_{NS}(t) + K_1(\theta^*_{NS}(t) - \theta_{NS}(t)) + \\ + K_2(\dot{\theta}^*_{NS}(t) - \dot{\theta}_{NS}(t)) \quad (1)$$

where $\ddot{\theta}^*_{NS}(t)$, $\dot{\theta}^*_{NS}(t)$ and $\theta^*_{NS}(t)$ represent the reference acceleration, angular velocity and angle of the swing leg. The control law (1) for an underactuated walking robot contains two free parameters, K_1 and K_2. These parameters must be set according to the physical parameters of the robot and the reference trajectories. We are currently verifying various approaches how to find these parameters, including one based on evolutionary algorithm shows in Fig 2.

The use of the evolutionary algorithm is included in the proposed methodology of walking robots control (Fig. 3). The results of the boundary value problem application are shown in Fig 4.

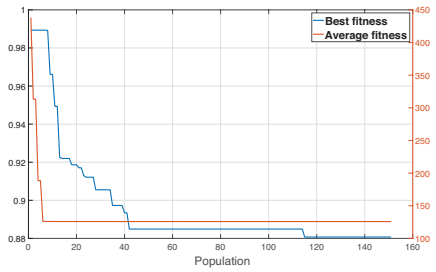


Figure 2. Fitness function of the evolutionary algorithm for K_1 and K_2

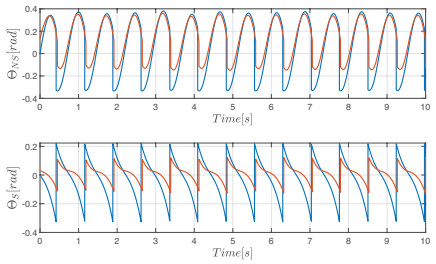


Figure 3. Reference trajectories tracking using the walking robot with PFL

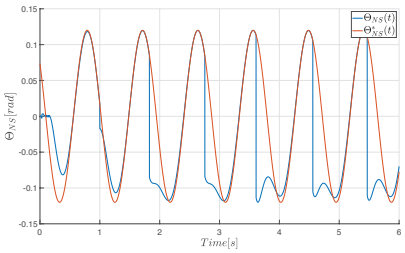


Figure 4. Walking robot with PFL using desired trajectories generated from BVP algorithm

The author's research group (Center of Modern Control Techniques and Industrial Informatics) participates in the research project *ALICE experiment at the CERN LHC: The study of strongly interacting matter under extreme conditions* [11], which involves every member of the group. As part of his research activities in CERN, the author attended two scientific internships over the last year. During these scientific internships, a framework was designed to create applications in WinCC OA which communicate within the Alfred architecture to control the ALICE's detectors.

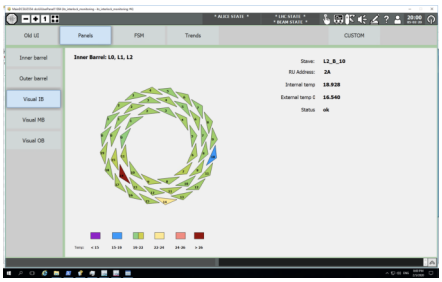


Figure 5. Application for monitoring inner barrel of the ITS detector

IV. Future Research Steps and Conclusion

This article summarizes the author's research activity over the last year. The main contribution is the proposed methodology which is used in modelling and control of the walking robots. The aim of the thesis has been to design the methodology of generating mathematical models of hybrid walking robots with different configurations which will enable the verification of designed control methods using suitable control structures.

Acknowledgment

This work has been supported by grant KEGA Implementation of research results in the area of modeling and simulation of cyber-physical systems into the teaching process - development of modern university textbooks – 072TUKE-4/2018 (100%).

References

[1] A. Goswami, B. Thuilot, and B. Espiau, "Compass-like biped robot part i: Stability and bifurcation of passive gaits," Ph.D. dissertation, INRIA, 1996.

[2] L. Koska, "A survey of approaches for modeling and control of effective walking robots," in *SCYR 2018*. TU, 2018, pp. 16–19.

[3] Y. Liu and H. Yu, "A survey of underactuated mechanical systems," vol. 7, no. 7, p. 921–935, 2013.

[4] T. McGeer *et al.*, "Passive dynamic walking," *I. J. Robotic Res.*, vol. 9, no. 2, pp. 62–82, 1990.

[5] L. Koska, S. Jadlovská, D. Vošček, and A. Jadlovská, "Inverted pendulum with linear synchronous motor swing up using boundary value problem," *Acta Polytechnica*, vol. 59, no. 5, pp. 458–466, 2019.

[6] L. Tedrake, Russel, *Underactuated Robotics: Learning, Planning, and Control for Efficient and Agile Machines*, 2009.

[7] S. Jadlovská, L. Koska, and M. Kentoš, "Matlab-based tools for modelling and control of underactuated mechanical systems," *Transactions on Electrical Engineering*, vol. 6, no. 3, 2017.

[8] L. Koska, "Modeling and control for walking robots using hybrid systems," *SCYR 2020*, pp. 134–135, 2020.

[9] A. Goswami, B. Espiau, and A. Keramane, "Limit cycles in a passive compass gait biped and passivity-mimicking control laws," *Autonomous Robots*, vol. 4, no. 3, pp. 273–286, 1997.

[10] L. Koska, "Modeling, analysis and simulation of the hybrid models of the walking robots," Ph.D. dissertation, Technical University of Košice, 3 2019, dissertation prospectus.

[11] P. Chochula, L. Jirden, A. Augustinus, G. De Cataldo, C. Torcato, P. Rosinsky, L. Wallet, M. Boccioli, and L. Cardoso, "The alice detector control system," *IEEE Transactions on Nuclear Science*, vol. 57, no. 2, pp. 472–478, 2010.

Simulation Model of a Small Hydropower Plant

[1]*Richard OLEXA (2nd year)*
Supervisor: [2]Pavol FEDOR

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

[1]richard.olexa@tuke.sk, [2]pavol.fedor@tuke.sk

Abstract – Nowadays, the vision of being environmental friendly, or green is getting more and more attention almost all over the world. Of course, there are plenty of ways how to improve our impact on the nature. Thus, probably everyone should think about, how he or she could help the society with this task, but, on the other side, naturally, not everyone is a scientist. The scientists have a bit more options and knowledge, that should be used in a proper way to enhance this alarming need of humans of being less harmful and more responsible to the nature. We considered using small hydropower plants could be one of the widely spread renewable sources of the energy in the future, but only if its structure and controllers are efficiently designed. For this purpose, the design of the simulation model of such a system was chosen as our first task. This article presents a summary of the work held during the last months on the topic of creating a simulation model of a small hydropower plant (SHP).

***Keywords*—Hydraulic turbines, hydropower systems, modeling and controlling hydropower plants, small hydropower plant, hydraulic turbine's efficiency, asynchronous generator**

I. Introduction

Renewable energy sources; being environmental friendly; waste recycling; carbon footprint; and many other terms are nowadays widely used and spread around the world. Moreover, it is no more only the topic of the politicians, but the ordinary people are very interested in this field, too. Therefore, the pressure on the researchers in various fields is now being induced to help the society solve maybe so far, the biggest problem of the human race.

Thus, we chose small hydropower plants as our subject of the research, mainly to find a proper simulation model for such a system to provide the potential investor an easier and cheaper way of designing and tuning of the control mechanism even before the construction work begins. Such a solution could make small hydropower plants more attractive to investors and thus, it could help to spread these systems and make the share of the electricity produced by renewable energy sources a bit bigger.

Therefore, the first initial aim was clear, we had to analyze all already published methods of modeling hydropower plants and then to find a proper way of combining them into one simulation model that should meet the needed requirements.

II. Initial Status in Modeling of Hydropower Plants

There are many scientific papers dealing with the topic of modeling and controlling hydropower plants. However, the vast majority of them [1]-[2]-[3] are mainly interested in studies of big scale hydropower plants, where a lot of phenomenon and technical parts have to be included in the model, e.g. in small hydropower plants, there are usually not as long penstocks as in the big ones are and thus, the water-hammer phenomenon could be neglected.

On the other hand, a lot of studies had been done via proposing a simple version of small hydropower plant model, i.e. designing it by using linearized transfer functions for each of the simulated subsystems. [1]-[2]-[3]-[5]-[6]-[7]

Moreover, almost each of the published simulation models is modeled with electric generator subsystem, while considering a use of a synchronous machine for this purpose. [4]-[8]-[9]-[10]-[12] Thus, not many authors are investigating an option of asynchronous generator as an alternative.

III. Proposed Scheme of Small Hydropower Plant's Simulation Model

As we discuss and assume in the article [11], the simulation model of a small hydropower plant usually consists of four main parts, or sub-systems, plus a grid model. The interconnections and main state variables are depicted in Figure 1.

Dividing the model into four sub-parts meant the necessity of studies and a design of each one, as a separate model. Brief description of some of these parts is mentioned below, while more details are summarized in the papers of [11]-[13]-[14]. As shown in the Figure 1, control system, or so called governor is the one responsible for keeping the operation of a simulated SHP within the pre-set ranges of controlled parameters. There are various types of designing and tuning PID controller [1]-[2]-[3]-[11], but during the last year, unfortunately, mainly because of the pandemic, we were not able to measure and acquire necessary data on the real system, needed for a proper design and governor tuning process. Therefore, we focused mainly on other parts of the proposed SHP's simulation model structure.

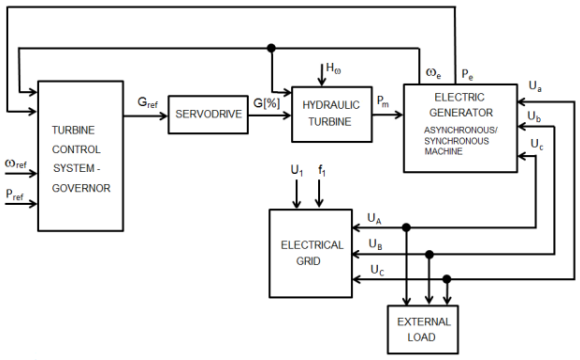


Figure 1 Block diagram of a Small Hydropower Plant [13]

Considering so far published papers and literature on this topic, we propose a simulation scheme of a servo-drive as a second subsystem that serves as an actuator for controlling the hydropower flow to the hydraulic turbine [11]. It is modeled by the second order dynamic system, which state variables represent the opening speed of blades of the inlet flap into the turbine and the position of opening of these flaps [11], as depicted in the Figure 2.

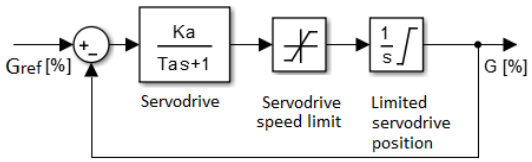


Figure 2 Block diagram of the Servodrive subsystem [11]

The before-mentioned paper describes the other parts of the SHP model, as well, where by using the parameters listed in the article [11], the below stated results (shown in the Figure 3) were obtained.

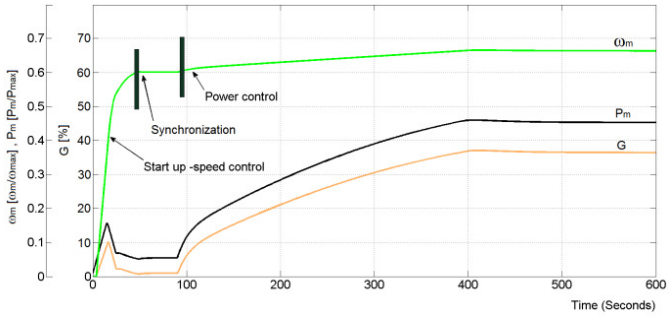


Figure 3 Simulation of the turbine's start up, synchronization (phasing) and its power control [11]

"*At time t = 0, the turbine starts to run to the grid frequency without connecting the generator to the grid. After reaching the speed corresponding to the grid frequency, the generator is phased (synchronized), i.e. it is connected to the power grid. After the transient state has been stabilized at t = 80 s, the "governor" using the logic signal /P, switches the mode from speed control to power control. At time t = 400 s the turbine is in operation. The power is stabilized at a standard value of Pm = 0.36 which corresponds to the nominal turbine power.*" [11]

According to the difficulties that came with all the restrictions meaning us not having an opportunity to measure the necessary parameters of a real SHP for model tuning purposes, we decided to aim at the way of representing of the hydraulic turbine within the simulation model. [13] An overview of currently used methods and a new proposed way of modeling hydraulic turbine's efficiency was described and published in the paper [14]. The reason for starting a research on a brand-new approach, i.e. fuzzy approach of hydraulic turbine's efficiency modeling was mainly supported by the fact that usually used methods are based on a calculated constant, or so called no-load flow, that makes it not very accurate, especially in case of present progressive wearing of the turbine's blades, or its runner (based on the turbine's type).

Figure 4 and 5 show the difference in the simulated efficiencies of hydraulic turbine based on the degree of wearing (the A_{blades} parameter) of its runner/blades. Such a difference could play very important role in proper SHP representation, and in its controller tuning process, respectively.

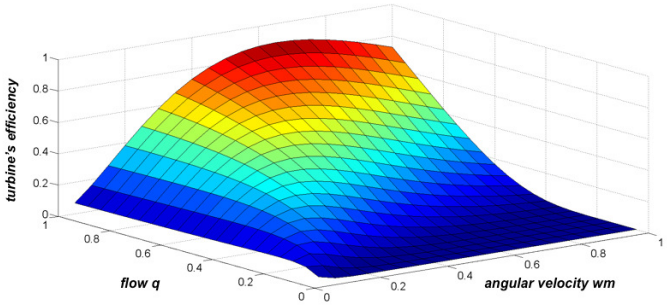


Figure 4 The efficiency of the hydraulic turbine (R_{blades} = 1.9m and A_{blades} = 9m^2) – Maximum turbine's efficiency is approx. 88% [14]

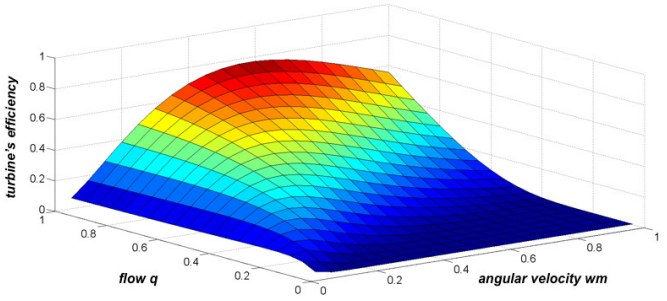


Figure 5 The efficiency of the hydraulic turbine (R_{blades} = 1.9m and A_{blades} = 8m^2) – Maximum turbine's efficiency is approx. 77% [14]

Currently, our research is focused on investigation of modeling an asynchronous generator as a part of SHP, its potential advantages and disadvantages, as well as on its possibility of being used in an islanding operation form. All of these aspects have not been widely investigated in comparison to the similar use of synchronous machine. Therefore, based on the scientific papers [15]-[16]-[17]-[18], we aim to study this topic deeper and complete a complex simulation model of SHP with asynchronous generator.

The next step will be then a data acquisition (on the real SHP) – it means measuring of all the needed parameters of the SHP, its characteristics, etc. Verification of the proposed simulation model based on the measured data must be done following the before-stated phase and having a verified simulation model gives us an option to design a proper controlling structure, as well as to determine its parameters and to potentially verify the setup obtained from the model in the real operation, on a real SHP system.

ACKNOWLEDGMENT

This work was supported by the Faculty of Electrical Engineering and Informatics, Technical University of Kosice,

Slovakia, under the Grant FEI-2020-64.

REFERENCES

[1] IEEE Working Group., "Hydraulic turbine and turbine control models for system dynamic studies," in IEEE Trans on Power Syst., 1992, 7, pp.167–79.

[2] H.J. Wagner, J. Mathur, Introduction to Hydro Energy Systems: Basics, Technology and Operation. Heidelberg : Springer, 2011. ISBN 978-3-642-20708-2.

[3] G. Munoz-Hernandez, S. Mansoor, D. Jones, Modelling and Controlling Hydropower Plants. 1st Edition. London : Springer London Ltd, 2013. s. 299. ISBN: 978-1-4471-6221-6.

[4] D.G. Ramey, J. W. Skooglund, "Detailed hydro governor representation for system stability studies," in *IEEE Trans on Power Apparatus and Systems*, 1970, 89, pp.106–12.

[5] J. Tiwari, et al., "Modelling and Simulation of Hydro Power Plant using MATLAB & WatPro 3.0.," in *Intelligent Systems and Applications*, 2015, 08, pp. 1-8. DOI: 10.5815/ijisa.2015.08.01.

[6] G. Singh, D.S. Chauhan, D. S., "Simulation and Modeling of Hydro Power Plant to Study Time Response during Different Gate States," in *(IJAEST) International Journal of Advanced Engineering Sciences And Technologies*, 2011, Vol. No. 10, Issue No. 1, pp. 042 – 047.

[7] R.A. Naghizadeh, S. Jazebi and B. Vahidi, "Modelling Hydro Power Plants and Tuning Hydro Governors as an Educational Guideline," in *International Review on Modelling and Simulations (I.RE.MO.S)*, 2012, Vol. 5, No. 4., pp.780-1790.

[8] A. Acakpovi, E. B. Hagan and F. X. Fifatin, "Review of Hydropower Plant Models," in *International Journal of Computer Applications*, vol. 18, December 2014, pp. 33-38.

[9] A. A. Usman, A. Abubakar and R. A. Abdulkadir, "Modelling and Simulation of Micro Hydro Power Plant Using Matlab Simulink," in *Proc. 2nd International Conference on Science, Technology and Management*, New Delhi: ICSTM, 2015, pp. 1121-1133.

[10] M. Sattouf, "Simulation Model of Hydro Power Plant Using Matlab/Simulink," in *International Journal of Engineering Research and Applications*, vol. 1, Brno: IJERA, 2014, pp. 295-301. ISSN: 2248-962

[11] P. Fedor, D. Perduková, R. Olexa, "Modelling a Small Hydropower Plant". In *Proc. Energetika 2019*, Stará Lesná, pp. 352-356. ISBN: 978-80-553-3324-3

[12] C. Jaliu, I. Visa, D. "Diaconescu, Dynamic Model of a Small Hydropower Plan", In *Proc. 12th International Conference on Optimization of Electrical and Electronic Equipment*, 2010.

[13] R. Olexa, P. Fedor, D. Perduková, "Model of hydraulic turbine taking into account the impact of efficiency". In *Mathematical modeling: international scientific journal*, Sofia (Bulharsko), 2020, pp. 64-67. ISSN 2535-0986

[14] R. Olexa, D. Perduková, P. Fedor. "Fuzzy Approach of Modeling a Hydraulic Turbine Efficicency". In MM Science Journal, Prague, November 2020, pp. 4086-4092. ISSN 1803-1269

[15] M.H. Haque. "Characteristics of a stand-alone induction generator in small hydroelectric plants." In. Proc. *AUPEC '08: Australasian Universities Power Engineering Conference*, 2008, pp. 1-6

[16] J.B. Ekanayake. "Induction generators for small hydro schemes". In *IEE Power Engineering Journal*, April 2002, pp. 61-67

[17] N.P.A. Smith. "Induction generators for stand alone micro-hydro systems". In *Proc. of the 1996 International Conference on Power Electronics, Drives and Energy Systems for Industrial Growth*, 1996, vol. 2, pp. 669-673

[18] C.P. Ion, C. Miranescu. "Autonomous micro hydro power plant with induction generator". In *Journal of Renewable Energy*, 2011, pp. 2259-2267

Overview of Modeling and Simulation Methods for Mechanical Systems

[1]*Martin DURKÁČ (1st year),*
Supervisor: [2]*Ján JADLOVSKÝ*

[1,2]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]martin.durkac@tuke.sk, [2]jan.jadlovsky@tuke.sk

***Abstract*—This paper describes different approaches to mechanical systems modeling with the focus on the new modern approach called physical modeling. Two selected case studies are presented, where an underactuated benchmark system is modeled via physical modeling and compared with the results of analytical identification. Research in the area of physical modeling is the main topic of my dissertation thesis.**

***Keywords*—physical modeling, analytical modeling, simulation tools, simulation, mechanical systems**

I. Introduction

Nowadays, mechanical systems modeling is studied by many research institutes and universities in the world. Modeling of such systems can be done either by using the analytical approach i.e. computing the equations of motion using Euler-Lagrange method [1], Kane method [2] or via physical modeling, enabled i.e by the Simscape toolbox in MATLAB. This toolbox allows the user to create a model using the Simscape Multibody library blocks or with using of Computer Aided Design (CAD). For a significant portion of physical modeling examples developed in MATLAB, SimMechanics toolbox (predecessor of Simscape) was used [3].

Researchers from the Center of Modern Control Techniques and Industrial Informatics at the Technical University of Košice have analyzed different approaches to modeling of underactuated nonlinear systems. Recently, the potential of Simscape/SimMechanics for creating simulation models has been increasingly explored, with the resulting models often compared to those obtained via analytical methods. This approach is inspired by the comparison analysis performed by research groups worldwide and applied to benchmark systems, such as the cart-pole system [3], double inverted pendulum [4], Spatial 6-DOF Parallel Robots [2], two-link robot [5] and inverted rotary pendulum [1]. Physical modeling has also been used to create a prototype or a digital twin for a real laboratory model, or to develop a way to obtain better visualisation of complex systems' behavior as a result of various control strategies. Applications include: virtual excavator [6], robotic arm [7], double suspension [8], self-balancing platform [9], robotic platform [10] and two wheel self-balanced robot [11].

II. Overview of mechanical systems

The mechanical system uses force to perform its task. The force acting on the mechanical system changes its movement along a specific trajectory, which depends on the mechanical arrangement of the bodies in the system. The system contains an energy source that converts energy to create force and movement. Sensors are used to obtain the response of the mechanical system and track the generated motion [12] [13].

In general, a body has six degrees of freedom, three of which are sliding and three rotating. The degree of freedom of a mechanical system is determined in physics as the number of parameters that describe the state of a physical body and are independent of each other. If the system consists of two particles and each particle has three degrees of freedom, the system as a whole has six degrees of freedom. We reduce the number of degrees of freedom by using articulated joints [14]. As in [15], a mechanical system consisting of N free particles will have 3N degrees of freedom. Each constraint of system reduces the number of degrees of freedom by one. If a system of N particles is acted upon by M constraints, the number of necessary generalized coordinates needed to describe the motion is

$$s = 3N - M, \tag{1}$$

According to [16], mechanical systems are divided into two groups: fully-actuated and underactuated. Fully actuated systems contain exactly as many inputs to the system as the degrees of freedom. Any system that has fewer inputs to the system than its number of degrees of freedom belongs to the underactuated systems. In addition to the two groups in [16], we also know a third group of systems called non-actuated, which do not contain any actuator, and thus the degrees of freedom are not controlled.

III. Methods for modeling of mechanical systems

Several approaches to mechanical systems modeling are recognized. The basic ones include analytical identification, experimental identification and physical modeling. In this section, I will briefly describe the first two named and I will describe the physical modeling, which I will focus on in my dissertation thesis.

A. Analytical identification

The analytical identification method of model derivation is based on the knowledge of underlying physical laws, while often neglecting real physical properties. It relies on knowledge of the structure of the object, material points and the physical forces exerted on interconnected material points. We assume that the position of these material points does not change over

time. The bodies are often considered to be infinitely thin, rigid and non-deformable [17] [18] [19].

The basic approaches to analytical identification include the application of Newton's and Euler's laws [20]. Moreover, Lagrange's equations of the second kind are among the most used and the most popular methods of deriving the mathematical model [13] [21]. The advantage of these equations includes the use of generalized coordinates as well as suitably selected system parameters, in particular distance and angle parameters. When deriving equations by the Lagrange method, it is necessary to proceed directly from the d'Alembert principle, which describes virtual works. We define the Lagrange function according to [22] as the difference between kinetic and potential energy. Two types of forces act on our resulting system. Conservative forces maintain the energy balance given by the kinetic and potential energy of the system. Non-conservative forces, on the other hand, interact with conservative forces and thus upset the energy balance. The disturbance occurs mainly by the action of a speed-dependent damping force.

The use of classical modeling methods using Lagrange, Hamilton, Newton leads to the same description of the same system. Complicated and often lengthy derivation of equations for more complex systems using the mentioned methods has led to the creation of a new method for deriving equations of motion. Kane's method [2] provides a simple framework for deriving simpler equations of motion by offering a physical overview of the system. Since the method does not use the relationships based on virtual work, it relies on the relationships defined at the basic mathematical level. It also allows to work with non-holonomic systems without the need to introduce and eliminate Lagrange multipliers.

B. Experimental identification

The experimental identification method relies on the existence of a real, physical model of the system. Experiments are gradually performed on the physical model in order to obtain values on the basis of which we can identify the system without knowing the laws of physics and dynamics. In these experiments, we let different signals of different types to enter the system and measure the individual outputs of system [23].

These signals can act on the system as actuators and include step and ramp, oscillating signals and pulse signal. Other inputs that affect the real system include measurable and non-measurable errors. However, these inputs cannot be influenced, although it is possible to measure external errors that fall under measurable errors. Since they can only be measured, we cannot influence them under normal conditions [24]. After measuring the data from the sensors and storing all the values, the last part in the process of experimental identification of the system occurs - obtaining the model. For the correct approximation of the real system, it is necessary to carefully choose the method and process the obtained data. After successful identification, it is advisable to verify the model. [24] [25]

C. Physical modeling

Physical modeling is based on the assumption that we do not have a real system for the method of experimental identification and modeling using analytical methods is too difficult to use for our system. Analytical identification methods suffer from the disadvantage that bodies that are part of a system are often represented in an idealized way that does not entirely describe to the physical system. The latter is composed of components of various shapes created from several types of materials to ensure the stability and safety of the entire physical structure. Therefore, if we do not have a model created so that we can use experimental identification, physical modeling becomes the most suitable choice. [26]

One of the first steps in physical modeling is to design the physical structure of the system. We can quickly and clearly create a physical structure using CAD tools. Modern CAD tools can help us in the design of these systems through various simulations of the mechanical load on the objects we create. The design itself should be consulted with experts in the field of mechanical engineering. [4]

Based on the physical properties of materials, the objects are assigned the mass and position of the center of gravity. The individual components of the structure must be interconnected by means of articulated connections. Within physical modeling, we distinguish several types of joints, while the basic ones include the revolute joint, the prismatic joint, the planar joint, the gimbal joint and the 6-DOF joint.

The revolute joint is one of the simplest types of joint that allows rotation within one axis. The principle lies in connecting the two components with the holes by means of a cylindrical body. The revolute joint provides one free degree of rotation. The prismatic joint, like the revolute joint, contains only one translational degree of freedom. In the case of a prismatic joint, the static member is square in shape and the connected second member, which is also square, slides along it. Restriction on the shape of the object causes a limitation in degrees of freedom. The planar joint contains 3 degrees of freedom. The two degrees of freedom are translational and move the link along the X and Y axes. The third degree of freedom is rotation around the Z axis of the movable link. The gimbal joint is one of the articulated joints in which the initial coordinates of the two connected frames remain at the same point. Its 3 degrees of freedom allow the connected body to rotate around the X, Y and Z axes. We can find similar functionality in the case of a spherical articulation, but as there are problems with the limits to which it can rotate, a joint that does not have these limits. The 6-DOF joint contains six degrees of freedom: three translational ones allow the connected body to be moved along the X, Y and Z axes, three rotational ones allow the body to rotate around the X, Y and Z axes.

IV. Simulation tools

There are several simulation environments suitable for simulating mechanical systems.

MATLAB is undoubtedly one of the most sophisticated environments for systems analysis and design. We can find its application in machine learning, in signal and image processing, in computer vision applications, in control design and much more. MATLAB has its own language for the simplest possible implementation of computational operations using a matrix expression. Graphical tools enable fast and high-quality data visualization for subsequent analysis.

Simulink is an extension of MATLAB, which enables the creation of block diagrams for the simulation of multidomain problems. Simulink also provides customizable block libraries, importing functions from the MATLAB environment in the form of blocks, and advanced solutions for modeling and simulating linear and nonlinear dynamic systems.

Simscape is one of the toolboxes created for the use in the MATLAB/Simulink environment. It is used to create physical models of systems in the Simulink environment. It is one of the basic prerequisites for starting to create physical models. The individual blocks of the Simscape toolbox represent physical elements of the real world, such as pumps, motors and operational amplifiers.

Simscape Multibody is part of the Simscape toolbox, which provides 3D mechanical system modeling. We can model the system directly in the Simulink environment, or we can import a system created in a CAD tool. The model is formed by the correct interconnection of body blocks, articulated connections and various actuators. Simscape toolbox automatically computes the dynamics according to the physical connections between the blocks. The simulation also contains a direct 3D visualization of our system and after starting the simulation we can see the response of the system.

V. Comparison of different modeling methods

In this section, modeling of the Ball and Beam system is presented and different modeling approaches are compared. Ball and Beam system is modeled with two different versions: with the sliding ball without rolling and with the rolling ball. Both versions are modeled using physical and analytical modeling approaches. Schematic diagram of ball and beam model is shown in Fig. 1.

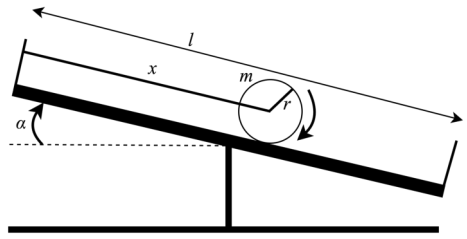


Fig. 1. Schematic diagram of Ball and Beam model

A. Ball and Beam without rolling

Ball and Beam system without rolling consists of a beam connected to World Frame and the ball sliding on the beam. The sliding movement is achieved using Prismatic Joint aligned with beam in distance equal of radius of ball itself. Physical model only contains two physical parts and one Joint. Block diagram of Ball and Beam model without rolling is shown in Fig. 2.

The analytical model of Ball and Beam system is derived from second Newton's law with respect to the inclined plane angle. The equation of motion is implemented to Simulink block diagram according to equation 2.

$$\ddot{x} = -\frac{b}{m}\dot{x} + g\sin\alpha, \tag{2}$$

B. Ball and Beam with rolling

Ball and Beam system with rolling consists of the same structure as previously mentioned Ball and Beam without rolling. The rolling movement is not achievable using basic blocks of Simscape Multibody library, thus Contact Forces library has to be implemented. The ball is connected with 6-DOF Joint to allow full range of movement possible. The

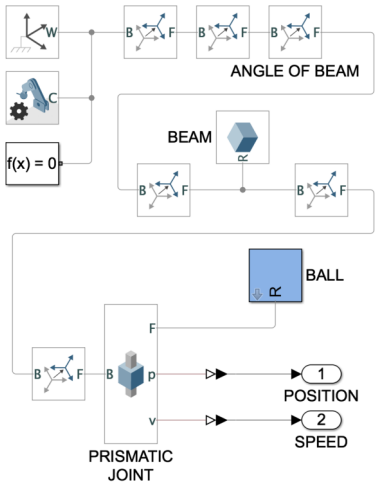


Fig. 2. Simscape model of Ball and Beam without rolling

block Sphere to Plane from Contact Forces library is connected directly to ball and plane itself. Block diagram of Ball and Beam system with rolling is shown in 3.

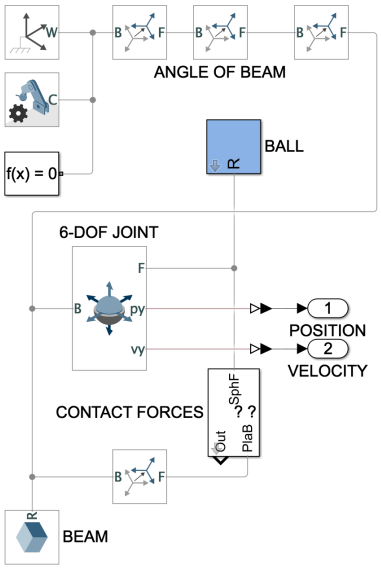


Fig. 3. Simscape model of Ball and Beam with rolling

Analytical model of Ball and Beam system is derived via Lagrange equations, where rolling movement with sliding movement of ball is taken into consideration. Simulink block diagram is implemented according to final equation 3.

$$mg\sin\alpha = -\left(m + \frac{J}{r^2}\right)\ddot{x} - b\dot{x}, \tag{3}$$

C. Comparison of dynamics

We next compare the dynamics from Ball and Beam model modeled without rolling V-A and with rolling V-B obtained via different methods. First of all, we compare models derived in V-A and also the physical model from V-B containing Contact Forces library with the deactivated rolling feature. Fig. 4 shows the comparison of Ball and Beam system without rolling.

The dynamics of Ball and Beam with rolling is shown in Fig. 5. This time only physical model derived using Simscape

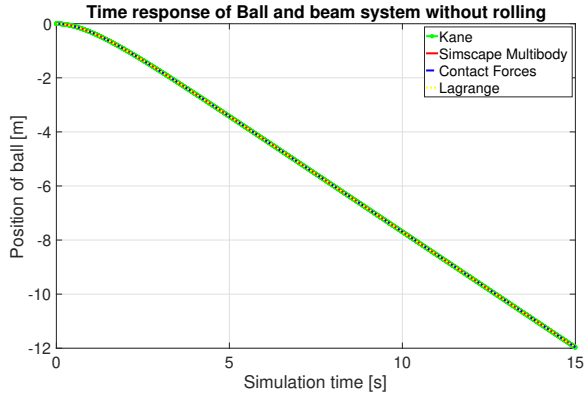


Fig. 4. Comparison of Ball and Beam without rolling

Multibody with Contant Forces library is used with two analytical models modeled via Euler-Lagrange and Kane method.

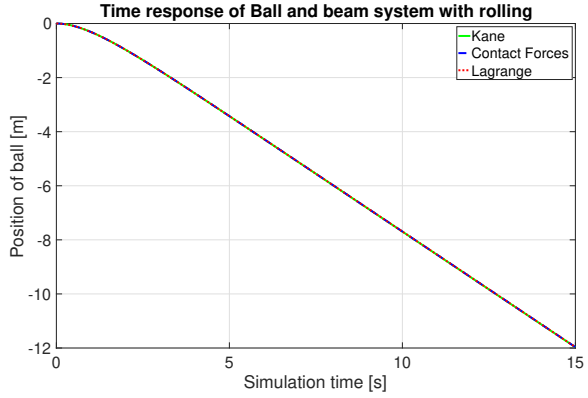


Fig. 5. Comparison of Ball and Beam with rolling

From Fig. 4 and 5 is obvious that response of both systems are identical for different modeling approaches.

VI. Conclusion

This paper provides a basic notion of mechanical systems modeling. We describe the approaches to modeling of mechanical systems, and using two case studies, we show how to model an example mechanical system using physical and analytical modeling. Because of the complexity of the more advanced systems, the considered modeling methods are demonstrated on a benchmark Ball on Beam model. We modeled the Ball and Beam in two versions: without the rolling ball and with the rolling ball.

In my dissertation thesis, I plan to focus on physical modeling of more complex mechanical systems, and on the subsequent design and implementation of suitable control algorithms. Another goal of my dissertation thesis is to create a virtual laboratory platform of mechanical systems represented as digital twins. The data provided by the platform will be stored in a database, which also allows to perform analyses and simulations later on.

The author is also a member of ALICE Collaboration in CERN ALICE experiment, which is included in the topics of his dissertation thesis.

Acknowledgment

This work has been supported by grant KEGA Implementation of research results in the area of modeling and simulation of cyber-physical systems into the teaching process - development of modern university textbooks – 072TUKE-4/2018.

References

[1] A. Kathpal and A. Singla, “Simmechanics™ based modeling, simulation and real-time control of rotary inverted pendulum,” in *2017 11th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, 2017, pp. 166–172.

[2] C. Yang, Q. Huang, Z. Ye, and J. Han, “Dynamic modeling of spatial 6-dof parallel robots using kane method for control purposes,” in *2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2. IEEE, 2010, pp. 180–183.

[3] M. Schlotter, “Multibody system simulation with simmechanics,” *University of Canterbury*, pp. 1–23, 2003.

[4] Y. Shaoqiang, L. Zhong, and L. Xingshan, “Modeling and simulation of robot based on matlab/simmechanics,” in *2008 27th Chinese Control Conference*. IEEE, 2008, pp. 161–165.

[5] T. Lepikhin, S. Pogozhev, and A. Vedyakova, “Main features and comparison of mathematical and physical modeling in matlab for two-linked robotic,” in *2014 2nd 2014 2nd International Conference on Emission Electronics (ICEE)*. IEEE, 2014, pp. 1–4.

[6] Q. Le, Y. Jeong, C. Nguyen, and S. Yang, “Development of a virtual excavator using simmechanics and simhydraulic,” *Journal of Drive and Control*, vol. 10, no. 1, pp. 29–36, 2013.

[7] R. U. V. Fedak, F. Durovsky, “Analysis of robotic system motion in simmechanics and matlab gui environment,” in *In MATLAB Applications for the Practical Engineer*, 2014.

[8] A. Tandel, A. Deshpande, S. Deshmukh, and K. Jagtap, “Modeling, analysis and pid controller implementation on double wishbone suspension using simmechanics and simulink,” *Procedia Engineering*, vol. 97, pp. 1274–1281, 2014.

[9] G. Madhumitha, R. Srividhya, J. Johnson, and D. Annamalai, “Physical modeling and control of self-balancing platform on a cart,” in *2016 International Conference on Robotics: Current Trends and Future Challenges (RCTFC)*. IEEE, 2016, pp. 1–6.

[10] J. Olaya, N. Pintor, O. F. Avilés, and J. Chaparro, “Analysis of 3 rps robotic platform motion in simscape and matlab gui environment,” *International Journal of Applied Engineering Research*, vol. 12, no. 8, pp. 1460–1468, 2017.

[11] S. Mohapatra, R. Srivastava, and R. Khera, “Implementation of a two wheel self-balanced robot using matlab simscape multibody,” in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. IEEE, 2019, pp. 1–3.

[12] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of physics*. John Wiley & Sons, 2013.

[13] H. Goldstein, C. Poole, and J. Safko, “Classical mechanics,” 2002.

[14] R. Grepl, *Modelování mechatronickỳch systém v Matlab SimMechanics*. BEN-technická literatura, 2007.

[15] P. Hamill, *A student’s guide to Lagrangians and Hamiltonians*. Cambridge University Press, 2014.

[16] R. Tedrake, “Fully-actuated vs underactuated systems,” *MIT OpenCourseWare*, vol. 6, pp. 1–8, 2018.

[17] F. E. Cellier and J. Greifeneder, *Continuous system modeling*. Springer Science & Business Media, 2013.

[18] A. Tornambe, “Modeling and control of impact in mechanical systems: theory and experimental results,” *IEEE Transactions on Automatic Control*, vol. 44, no. 2, pp. 294–309, 1999.

[19] P. A. Fishwick, *Handbook of dynamic system modeling*. CRC Press, 2007.

[20] M. D. Ardema, *Newton-Euler Dynamics*. Springer Science & Business Media, 2006.

[21] F. A. Potra and W. C. Rheinbold, “On the numerical solution of euler-lagrange equations,” *Journal of Structural Mechanics*, vol. 19, no. 1, pp. 1–18, 1991.

[22] P. Noskievič, “Modelování a identifikace systém,” 1999.

[23] L. Ljung, “System identification,” *Wiley encyclopedia of electrical and electronics engineering*, pp. 1–19, 1999.

[24] M. Fikar and J. Mikleš, *Identifikácia systémov*. STU BA, 1999.

[25] H. Garnier, L. Wang, and P. C. Young, “Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance,” in *Identification of continuous-time models from sampled data*. Springer, 2008, pp. 1–29.

[26] The MathWorks, Inc. "simscape™ user’s guide," feb. 15, 2020. [Online]. Available: https://ww2.mathworks.cn/help/pdf_doc/physmod/simscape/simscape_ug.pdf

Image stitching based on drone technologies

[1]*Peter ŠUĽAJ (4th year)*
Supervisor: [2]Stanislav MARCHEVSKÝ

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]peter.sulaj@tuke.sk, [2]stanislav.marchevsky@tuke.sk

***Abstract*— This paper is focused on working with devices supporting virtual reality and flying devices called drones. In collaboration with virtual reality is created by an aircraft system that should simplify 'Any activity' with wide spectrum. The usability of the system in different areas is given as the main meaning of this work.**

***Keywords*— drones, image stitching, multi-view video, , UAV, video streaming,**

I. INTRODUCTION

Utilizing today's technology and technical progress, people open up different horizons in a wide spectrum. One of these is also the use of a combination of virtual reality and a flying device that can be drones [1]. These devices can fully replace people who would have to risk their lives in any situation. Or an example use of simulated drone in virtual reality where we can see monuments, museums and other natural phenomena from our homes. As a good example, the Notre Dame in Paris was destroyed by fire. Creating a simulation and implementing an imaginary drone in this simulation of a virtual reality house would be able to faithfully bring this monument closer to us.

II.• DRONE COMBINATION WITH VR•

There are several projects that link VR kits with drones. One of these projects is Oculus Drone [3] associated with Ar.Drone-webflight [4]. However, these examples use only one particular VR kit and one specific drone type. This is confirmed by research within a framework that provides a generic character access method for drone control using a VR headset visualization is paramount.

The aim of the work was to find common methods for combining different types/brands of drones and to find solutions for communicating with different types of VR kits. New drone and VR kits are still on the market, so the system must be scalable with minimal effort. The modules created should not be modified for new models and the interface should be clear enough to be able to modify specific methods for the new type of drone. Other problems of this research include the incompatibility of VR kits. Many VR kits are still evolving and their software is updated with radical changes to the API.

VR accessories: Monitored head movements and 3D rendering: The system used for imaging combines two advances from smartphones: high-resolution displays and precision and technology advanced sensors. Instead of displaying information on an area such as a screen somewhere in the room in front of the user. These displays are located on the head or they are a part of the helmet. l see represent reality. When we starting work on this paper, Oculus Rift By far the most popular platform for Virtual Reality. Figure 1 shows interface preview for smartphone. Oculus Rift has passed various pre-production models, known as development kits (DK). DK1 became available in late 2012 and DK2 in mid-2014.

DK2 selected as the first supported VR platform in the development of DroVR.

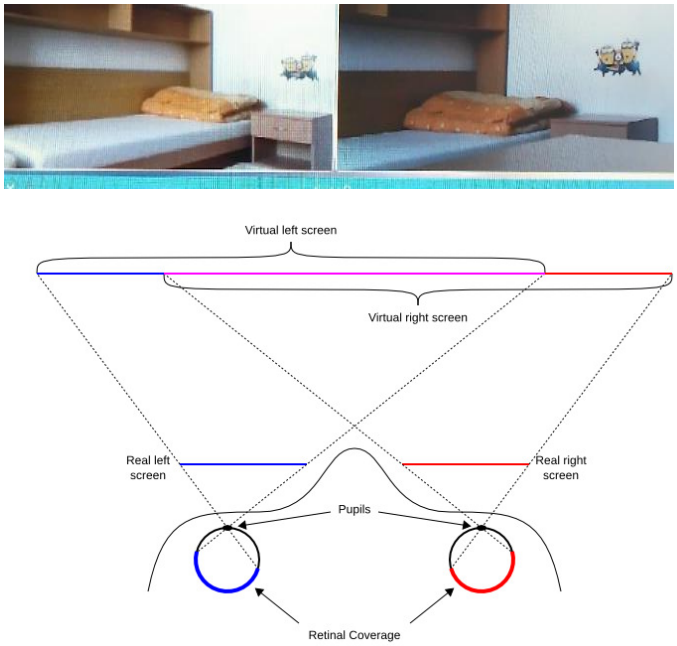


Figure 1 Show as VR concept for eyes

III. PROTOTYPE OF FRAMEWORK

The DroVR modular architecture is shown in the DroneAPI image and VRHeadsetAPI are the two main building blocks of the framework. These are generic modules that define which functions the modules must implement for specific one's drones and sets supporting VR. It should be clear that these components do not implement no functionality of a particular type of device. Their role is just to define and provide general information about the features that can be implemented for every new device. However, they will know which facilities have the prerequisites[2].

Another task of the architecture is to provide requirements for correct drone specifications and VR sets.For drones, DroneAPI is divided into two sub-components. On one side is the DroneControl component. This component defines the necessary methods physical controls for the drone, such as take-off and movement. On the other hand, NavDataReceiver

receives sensor information from a particular drone module. This can be battery level, altitude, video stream. For a VR network, the VRHeadsetAPI component consists of VRRendering as well as SensorDataReceiver. The first contains all the information needed to display on HMD for VRRenderingsComponent type VR set. For WebVR[2], this specific component provides snapshots from a camcorder via a WebSocket connection and serves as a REST API to enable Search all information using sensors. It's inevitable, as well as in WebVR, rendering is done in the browser. SensorDataReceiver receives all VR data. This information may be provided by type specific sensory component, as well as the REST API. Component usage example a sensor type specific to the engine would be, for example, reading the sensor data via specific SDK calls.

An example of using the REST API where reader data will be read in the browser and send it to the API endpoint. All communication between components is exclusively via asynchronous messages. Drone-Thus, the API defines messages as Drone Notifications.TakeOffRequestNotification. When it receives in this message it executes the takeOff method from the selected drone[4]. This method then runs the Command.TakeOff specification in that particular folder a drone that performs device-specific actions. The same paradigm is visible for VR set. VRHeadsetAPI defines generic messages such as EulerAnglesRequestNotification. If it receives messages, it executes the getEulerAngles method from connected VR headsets. LinkComponent is the link between the two building elements. This component decides what steps to take on the drone based on headset. The sensor data is received by the data sensors from the VRHeadsetAPI component and transmitted generic DroneAPI component actions. This increases the data flow rate. Because all sensor data is promoted through this module, it is also it is possible to add components that slip this data. For example, DroneEventProcessor can detect the battery level and can initiate if the drone battery is too low alert message. Since the sensor data also includes all camera images respectively. camcorders, we can do real-time video analysis. ObjectDetection component for example, it processes in real-time snapshots. Performs detection objects with a predefined classifier. When an object is detected by highlighting rectangle. One of the advantages of modular architecture is increased interoperability. If we want the frame to support a new device, either a drone or a VR set. it is necessary define a new device-specific. The only change you need to make in the system, the existing components register a new type of device. When connecting new device type must be specified. Such generic components know how to respond to earth type in specific method calls [5].

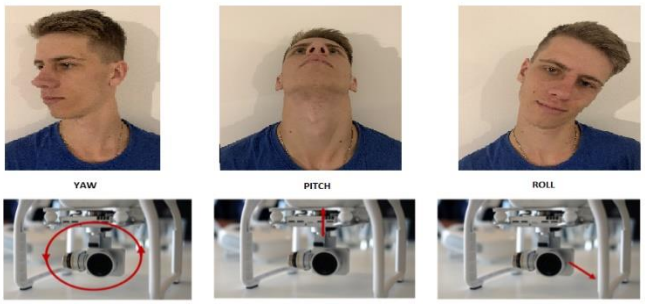


Figure 2 Gymbal control on drone

IV. Conclusion

Analyzing current VR and drone control solutions, we found that when combining different drones and different VR sets, we had to adjust different control methods and sensor data storage capabilities. We described the methodology for designing a generic framework prototype system and designed a modular architecture to address the issue of interoperability. When connecting different drones and VR sets, we used specific add-ons of manufacturers, where we can use API interface, which implements necessary add-ons for cooperation. If this is not defined, it is necessary to convert the UAV protocol. The difficulty of this process depends on the inspection methods used and the implementation possibilities of the drone manufacturer. We tested a prototype, which we can use for example in the border events as a smart city concept, where thanks to drones we can coordinate the emergency services very quickly. While expanding the system, we can extend the system with various modules such as real-time object detection, but these detections have a big impact on the sensitivity and refresh rate. in the future work is the possibility to implement and experimentally test different types of drones and their control in SWARM mode, which will allow us complete control of drones.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract no. “APVV-17-0208 - Resilient mobile networks for content delivery” and by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project KEGA 009TUKE-4/2019.

References

[1] Šul'aj, P., Haluška, R., Marchevský, S., Pulli, P., & Kramar, V. (2018, August). *UAV Management System for the Smart City.* In 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA) (pp. 119-124). IEEE.

[2] Muthana Najim Abdulleh, Salman Yussof, and Hothefa Shaker Jassim. *Comparative study of proactive, reactive and geographical* MANET routing protocols. Communications and Network, 7(02):125, 2015.

[3] V. Mersheeva and G. Friedrich. *Routing for continuous monitoring by multiple micro UAVs in disaster scenarios.* In Proc. European Conf. Articial Intelligence, 2012.

[4] HALUŠKA, Renát - ŠUĽAJ, Peter - OVSENÍK, Ľuboš - MARCHEVSKÝ, Stanislav - PAPAJ, Ján - DOBOŠ, Ľubomír: *Prediction of Received Optical Power for Switching Hybrid FSORF System / - 2020. In: Electronics. - Basel (Švajčiarsko)* : Multidisciplinary Digital Publishing Institute Roč. 9, č. 8 (2020), s. 1-14 [online]. - ISSN 2079-9292 (online)

[5] HALUŠKA, Renát - OVSENÍK, Ľuboš - ŠUĽAJ, Peter - KOŠČO, Milan: *Simulation of Object Recognition by Unmanned Areal Vehicles* / - 2019. In: Electrical Engineering and Informatics 10 : proceedings of the Faculty of Electrical Engineering and Informatics of the Technical University of Košice. - Košice (Slovensko) : Technická univerzita v Košiciach s. 371-374 [CD-ROM, online]. - ISBN 978-80-553-3342-7

Hybrid FSO/RF Systems and Cloud

[1]*Renát HALUŠKA (4th year)*
Supervisor: [2]Ľuboš OVSENÍK

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]renat.haluska@tuke.sk, [2]lubos.ovsenik@tuke.sk

***Abstract*— This article discusses the problem of optical space free communication (FSO), its use, and the conditions that affect it. Because of the vulnerability of the FSO transmission channel to weather conditions, such as fog, it is necessary to predict the Received Signal Strength Indicator (RSSI), and thus switch to a back-up RF (Radio Frequency) link using the decision tree method. The Microsoft Azure cloud service was chosen for this task, which is used for important data storage, backup, analysis and creation of a learning decision tree algorithm for a hybrid FSO/RF system.**

***Keywords*— FSO, hybrid system, Machine Learning studio, Microsoft Azure, radio communication**

I. Introduction

The demand for high bandwidth in short-lived metropolitan networks is increasing. In addition, the requirements of flexibility and cost-effectiveness of service provision create imbalances. This problem is often referred to as the "last kilometer bottleneck" [1].

The first most visible way to address the bandwidth shortage is fiber optics. Fiber is undoubtedly the most reliable means of optical communication, but digging, fiber storage costs and time to market are the most serious disadvantages of optical fiber. The second option is radio frequency (RF) communication [2]. This technology is advanced and often deployed today. Radio-based networks require huge investments to obtain spectrum licenses, but they cannot be compared to the transmission capacity of optical fibers. The third alternative is all copper line-based technologies, such as cable modems and DSL (Digital Subscriber Line). Although the copper infrastructure is available on a larger scale, the problem is bandwidth scalability [3]. The distance of copper per segment and performance is internally limited, so its potential to solve a close connection is also limited. The fourth alternative is communication in the free environment FSO (Free Space Optics). FSO represents the optimal solution in terms of technology, bandwidth scalability, deployment speed and cost efficiency [4].

Hybrid wireless space communication (FSO) and radio frequency (RF) communication is a way to provide reliable communication for critical real-time outdoor operation, as weather such as fog affects FSO much more than RF connections. The main condition for using FSO is that the receiver must be in direct line of sight with the transmitter [5].

II. Hybrid FSO/RF System

Free Space Optics (FSO) communication involves the transmission, absorption and scattering of light through the Earth's atmosphere. The atmosphere interacts with light due to the composition of the atmosphere, which normally consists of molecules of various gases and small suspended particles called aerosols

Hybrid Wireless Free Space Communication (FSO) and Radio Frequency Communication (RF) is a way to ensure reliable communication for critical real-time outdoor traffic as weather like fog affects FSO much more than RF connections [4]. The main condition for using the FSO is that the receiver must be in line of sight with the transmitter [5].

Due to different carrier frequencies, FSO and RF lines have different channels. The short wavelengths used in FSO links are severely affected by fog scattering and atmospheric scintillation, while RFs are particularly sensitive to rain scattering. Using both systems to overcome these problems can lead to increased reliability and speed [6].

When configuring hard FSO line switching, the transmitter and receiver together select either the FSO or the RF channel for data transmission. A high-speed FSO connection is only selected if the channel conditions allow reliable communication, otherwise all data is sent over the RF channel. The transmitter and receiver must be coordinated through feedback to select the correct channel for transmission. The key disadvantage of hard switching is that whenever one connection is idle. In practice, scintillation or loss in the sensitive FSO channel may cause its rare selection. When the RF connection is selected, the capacity of the FSO connection channel is lost. Figure 1 shows possible configuration of the FSO / RF hybrid connection [7].

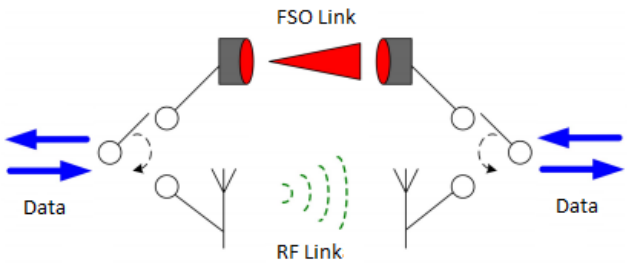


Fig. 1 Hybrid FSO / RF line

III. Prediction RSSI Parameter in Cloud

Machine learning techniques are widely used today for many other tasks. Different types of data related to different methods. The role of machine learning in FSO / RF hybrid systems is to predict the Received Signal Strength Indicator (RSSI) for which they have weather conditions [8]. To maximize the accuracy of the predictions, a form of machine learning and a transition to a decision tree that is designed to practice more complex methods is suggested. The transition enhancement is

useful for predictive models that analyze organized data and categorical data. Decision trees are used to predict RSSI classification [9].

Microsoft Azure Machine Learning, or Azure ML for short, is a complete cloud service. It is accessible via any web browser. This means it's not necessary have to buy any high-performance hardware. It is therefore a fully managed cloud service that allows analysts, data scientists and developers to create, test and deploy predictive analytics in their applications or in stand-alone analysis. Creating applications that use machine learning will turn into a service that allows you to visually build a learning model. Azure ML also makes it easy to get an overview of large data files and brings all the benefits of the cloud by integrating deep learning tools such as Hadoop.

Before performing the prediction, the algorithm must be trained with examples of data or a data set where the target value is known. After training the model, it is possible to make a prediction of the received optical power (RSSI) with the trained model. Usually, when a model has to be built from a given data set, the data set is divided into two sets, the first is used as the training set and the second as the test set. When a model is trained with data from a training set, we can use the test data set to find out how the model is doing, i.e. how many errors it has. After training the model, you can use the data from the test set for prediction or scoring. Element values are used for scoring and then the target value is predicted. During the evaluation, we take the scored value and the actual value, which is known by dividing the data set into a training and test set.

The input and training set is split to create a decision tree smaller subsets that gradually characterize the values of the output variables. The training matrix of the input variables for the FSO/RF system has the following structure:

$$P = \begin{bmatrix} P_1 & T_1 & G_1 & V_n & H_n & \cdots & W_1 \\ P_2 & T_2 & G_2 & V_n & H_n & \cdots & W_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_n & T_n & G_n & V_n & H_n & \cdots & W_n \end{bmatrix} \quad (1)$$

where P is barometric pressure (hPa), H represents ambient humidity (%), T is air temperature (°C), wind speed W (m/s), G represents concentration Airborne particulate matter (mg/m3) and visibility V (m). The matrix of the output variable y (target), which represents the received optical power of RSSI, is interpreted as follows:

$$y = [y_{RSSI.1} \quad y_{RSSI.2} \quad \cdots \quad y_{RSSI.n}]^T \quad (2)$$

By creating the variables x and y by taking them from the data set and using a function in the custom Python script of Azure ML Studio, we divide the data into training and test sets. A test size of 0.2 indicates that we used 20% of the data for testing. For the output train_test_split we get the values *X_train*, *X_test*, *y_train* and *y_test*. The complexity parameter used for pruning with minimum cost complexity was chosen *ccp_alpha* = 0.0. The function of measuring the quality of the distribution was chosen *mse* (mean square error). The minimum number of samples needed to split the internal node *min_sample_split* was chosen to be 10. The minimum number of samples that must be in the leaf node *min_sample_leaf* was selected to be 2. The maximum depth of the *max_depth* tree was derived from the *mse* value from size 2 to size 100. Best the value of the *max_depth* parameter was set to 5 for *mse* = 0.29833 (Fig. 2).

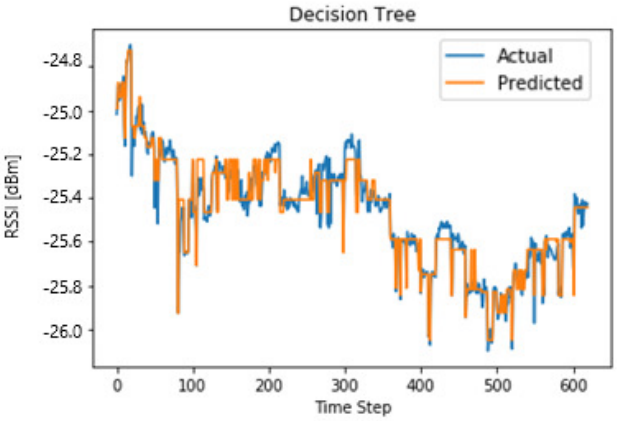


Fig. 2 measured and predicted data from the input matrix

IV. Conclusion

Due to the high impact of weather on FSO transmission, a backup RF line is required, which is less prone to, but much slower. For this reason, a hybrid FSO/RF line is required, which can be switched to the hard or soft switching principle. To create such a system, it is necessary to analyze the conditions and thus obtain the characteristics of the variables that affect the transmission channel. Appropriate RSSI prediction, using machine learning methods, makes it possible to quickly switch over communications over the RF line and back to a faster FSO line. In this way, it is possible to increase the efficiency of communication via the hybrid FSO/RF system.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract no. "APVV-17-0208 - Resilient mobile networks for content delivery" and by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the research project KEGA 009TUKE-4/2019.

References

[1] V. Brazda, O. Fiser, and J. Svoboda, "FSO and radio link attenuation: meteorological models verified by experiment," in *Free-Space and Atmospheric Laser Communications XI*, A. K. Majumdar and C. C. Davis, Eds. SPIE-Intl Soc Optical Eng, 89 – 104. sep 2011.

[2] P. Singal, S. Rai, R. Punia, and D. Kashyap, "Comparison of different transmitters using 1550nm and 10000nm in FSO communication systems," *International Journal of Computer Science and Information Technology*, vol. 7, no. 3, pp. 107–113, jun 2015

[3] R. Gupta and P. Singh, "Hybrid fso - rf system: A solution to atmospheric turbulences in long haul communication," *International Journal of Scientific Engineering Research*, vol. 5, no. 11, 103-108 nov 2014.

[4] J. Tóth, "High availability and reliability in wireless optics using data analytics techniques" - 2017.In: *SCYR 2017*. - Košice : TU, 2017 S. 118-119. - ISBN 978-80-553-3162-1

[5] T. Hastie and J. M. Chambers. "Statistical models." *Statistical Models in S*. Routledge, 2017. 13-44.

[6] H. Gavin. *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd, 2017. 130-146

[7] L. Rokach and O. Maimon, "Top-Down Induction of Decision Trees Classifiers—A Survey," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 4, pp. 476–487, 2005.

[8] M. Brzoska, "Modellierung zwischen „overfitting' und „underfitting'," *Rationale Entscheidungen unter Unsicherheit*, pp. 152–156, 2018.

[9] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technology*, vol. 25, pp. 326–332, 2016.

Smart Grid and Microgrid in an Industrial Zone

[1]*Rikin Tailor (1st year)*
Supervisor: [2]Ľubomír Beňa

[1,2]Dept. of Electrical Power Engineering, FEI TU of Košice, Slovak Republic

[1]rikin.jitendrakumar.tailor@tuke.sk, [2]lubomir.bena@tuke.sk

***Abstract* — It is certainly true that the 21st century is going to witnessing a revolutionary change in the variety of energy generation, transmission, distribution, and usage. In all energy, electrical energy has shown its significant dominance for widely usage energy. As electrical energy transforms into carbon-free and utilizing more renewable energy resources for generation, each sector in industries seeking solutions in form of electrical energy. This apparently creating a high global demand. These renewable energies are an intermittent source of energy that needed a smart network that automates all over infrastructure with efficient and reliable usage of energy. Hence, we come up with the Smart Grid and Microgrid technology solutions. According to heavy consumption, it is inevitable the importance of the involvement of Industries in Smart energy technology. In this paper, we are going to talk about Smart Grid and Micro Grid position in an Industrial Zone. Its challenges, status, and future direction.**

***Keywords*—Smart Grid, Micro Grid, Energy Management System, Cyber Security, Renewable Energy Resources, Distribution Generation.**

1. INTRODUCTION

Electrical energy, most versatile energy according to current and future scenarios. The transformation of electrical energy infrastructure has begun with the name Smart Grid. The Smart Grid concept is very old, but it becomes a one of the most dominant topics after 2010 as the demand for electricity increases rapidly. There is no exclusive definition of the Smart Grid, but all industries, companies, countries giving their own interpretation. However, the concept is the same, Smart Grid is built with a variety of elements such as renewable energy generation, smart transmission, smart distribution, distribution energy management, real time communication infrastructure, Internet of things, Cyber-Physical Systems, Distribution generation, distribution storage, autonomous control, islanding facility and most importantly self-healing process. The Smart Grid is very specific with time, location, and weather forecast, all these factors are very important for energy generation.

The industrial sector is the biggest consumer of electrical energy around 30% in some cases [1]. Table I. shows data of energy consumption of manufacturing technology in the US according to the manufacturing energy consumption survey - 2010 [2,3]. That means, the active participation of the Industrial sector is very important in Smart Grid technology to guarantee the energy system transformation. Even after the clean awareness about their influence, most studies on Smart Grids remain focused on residential/commercial consumers [4,5], and very few studies highlighting industrial consumers [4,6].

Table. I. Manufacturing technology energy consumption in US according to the manufacturing energy consumption survey-2010

Industry sector	**Total electricity used (10^6 kWh)**
Chemicals	207,107
Primary Metals	139,985
Paper	122,168
Food	78,003
Petroleum and Coal Products	60,149
Transportation Equipment	57,704
Plastics and Rubber Products	53,423
Nonmetallic Mineral Products	44,783
Fabricated Metal Products	42,238
Machinery	32,733
Wood Products	28,911
Computer and Electronic Products	27,542
Textile Mills	19,753
Beverage and Tobacco Products	17,562
Printing and Related Support	13,089
Electrical Equip., Appliances, and Components	12,870

The majority of Smart Grid features could be suitable for industrial power supply networks, even it would be easier in shaping as there will be a limited number of common coupling points. Industries are working with a high density of power and the distribution network is heavily loaded. Along with this, an industry would be a single entity for renewable/conventional generation and to connect with the grid. There would be no fluctuating power generation that makes deterioration in the quality of power supply in the system [7].

Reliable power supply with quality and energy efficiency is

one of the key factors for minimizing production cost. Many industries with sensitive production lines are more concerned about the rising power demand, fluctuation in generation with wind and solar and declination of stabilize power from the national electricity grid [8]. These events lead them to accomplish their own power generation and management system, which they can control on their own, the system more likely to be microgrid technology. This is happening generally in large industries that help to develop the quality of performance and record-breaking production [9]. It is notable that industries are already having pillars of Smart Grid technology. It is just needed some guidance for industrialist and plan to implement it. It would be more beneficial to start emphasizing Smart Grid and Microgrid technology in the Industrial zone.

Section II. presents challenges and the currents status of Smart Grid and microgrid development in industries. The solution of current problems and require directions discuss in section III.

2. Challenges and Status

Industries are eagerly seeking Renewable Energy Resources (RESs), Energy Storage Systems and Energy Management Systems (EMSs), because it is getting tough for them to obey equal regulation and evade expensive penalties with conventional fuel production and depend on outdated grid [10]. Smart Grid technologies can overcome these barriers; but with numerous development and technology expansions there could be certain barriers to adoption [11]. Challenges for the adoption of Smart Grids in industries are divided into two categories: A. Non-technical B. Technical.

A. Non-technical:

- A lack of clarity and interpretation created a reluctance to acceptance of Smart Grid technology. In one survey in Norway, representative shareholders agreed on this statement and indicate that at the time there was no proper definition of the subject Smart Grid [12].
- Government policies have been revised for the new equipment and infrastructure acceptance. This makes an extra effort for industries to adopt it.
- Smart Grid is still a new and emerging topic; hence it requires a lot of investment from the government for research and replacement of the current grid [12].
- Industries must redesign their internal network which maximally supports new Smart Grid devices, adoption of such new deportment and priorities is usually slow in the industrial environment [11].
- The higher capital cost of smart grid infrastructure along with the cost of new equipment, operating and security systems are another barrier for the industries.
- Industries invest in things that make them profit; Smart grid technology and Microgrid technology will not give instant benefits as per the investment, that discourages their moral to adopt, develop and implement.
- Here are eleven factors that have been identified by Zheng Ma and his colleagues:
  1. Awareness of Multiple Contexts
  2. Shared Support
  3. Return-on-investment.
  4. Ease of use
  5. Flexibility and dynamic pricing
  6. Liberalization and energy tariff structure
  7. Customer focus
  8. Solution integration
  9. Process improvement
  10. Service quality
  11. Company's green image

B. Technical:

- One of the most prominent drawbacks for initiation of Smart Grid infrastructure that mentioned in many of the literature is a deficiency of existing infrastructure. Which cannot cope up with the fast and advanced sensors and monitoring equipment [14,15].
- Reliable and efficient Renewable Generation equipment design, manufacturing and implement the location.
- Design and development of Electrical Storage Systems (ESSs)
- Design and development of smart inverter for interconnection with the main grid.
- Sensors, protection, monitoring and metering equipment design and development.
- Modern Grid design and development
- Smart Control Panels for control of street, region and city.
- Information and Communication Technology Challenges (CTs) [16]
- Big Data challenges for storage of large numbers of data [16]
- Cloud Computing Challenges [16]
- Cyber-Physical System (Security challenges) [16]
- Electric Vehicle and its charging infrastructure.
- Integration of all equipment, software and applications.

All this technology interconnection causing miscellaneous disturbances in power as sags, spikes, harmonics, etc. which is indispensable to solve before integration.

After comprehensive analyses in the last 10 years of paper regarding Smart Grid and Micro Grid technology, we have found that there has been a lot of research is done but still, technology is in the 1st and 2nd stage only. Most of the papers are concentrated on the primary and middle stages.

The 8 out of the 10 papers show the structure of study in a specific topic, some have developed simulations and possibilities or examine case studies and opportunities to further area before the task has been done [12,17].

Solar generation is the most dominating electrical generation alternative solution in renewable energy resources still, the most efficient solar panel converting 22.8% of total solar radiation into electrical energy, which significantly very less [18]. More ultra-modern work needs to be done for improvising efficiency.

The distributed storage system is an essential part of Microgrid system, there are many storage systems such as Flow battery [19], Concrete block and crane technology [20], Thermal Storage [21], compressed air technology [22], cryogenic energy storage system [23]; when it comes to electricity storage that even at the consumer side then Lithium-ion battery proven its significance among all other

storage systems. Lithium-ion has shown a major among the various batteries by comparing efficiency, size, max. rating, fast response, initial cost, and size. The downside of lithium-ion battery is life cycle, which is very less that makes it weak in the static storage system. The maximum expansion year for a lithium-ion battery to grow in all electrical sectors is 20-30 years. The biggest challenge to the lithium-ion battery is given by Prof. Donald Sadoway and his student David Bradwell via Ambri Technology (Liquid Metal battery) [24] and another competitive technology is static storage battery [25]. Apparently, both batteries are still in research and development.

One of the crucial issues in smart grid management is integration of distributed generation and forecast, which is clearly presented by William Sanchez-Huertas et al., which deals with optimization algorithms such as the Colony of Ants, Genetic Algorithm, and others for solving microgrid and smart grid integration problems. Smart grid integration problems are generation forecast, location, size, and configuration of RES. As one of the biggest problems in the problematics of smart grids is pointed out the management of the smart grids [26].

DNP3 (Distribution Network Protocol) and IEC 61850 are the protocols developed for the integration of the Smart grid to the existing power grid. DNP3 is designed for communication between the command center and a substation, whereas IEC 61850 is operating for communication within the substation [27]. In communication technology, another factor is data transmission technology. This data transmission technology used for transmitting user data and instruction messages between the various parts of the grid. Several data transmission technologies are in the market and under research. As an example, Fiber optic, Wireless, LPWAN (Low Power Wide Area Network), Zigbee, etc. Fiber optic is one of the finest data transmitting mediums with high-speed and high capacity, it is expensive material hence only utilized for long-distance data transmission requirements [28]. A 3G and 4G system is currently in use for the cellular networks, according to smart grid infrastructure data transfer rate ratio is slower in it. Future technology, 5G would be the most promising technology for data transmission in Smart Grid application as it has high data rates with a minimum message return time. A 5G technology is still evolving and in research requires some more years to adopt for Smart Grid system. This technology can be utilized for Wide Area Network and Neighbour Network [29,30]. Zigbee technology is designed to deploy in Home Area Network as it has a low bit rate, it has a greater advantage as it supports multiple network topologies [31]. A name itself suggest the type LPWAN; it is less power, low bit rate and long-distance data transmission topology. The two most frequently used protocols are LoRaWAN and Sigfox. LoRaWAN has the ability to send data between 2-15 km in a range from 0.3 to 50kbps of data rate [32] and Sigfox have up to 1000km range of data transfer with a lower data rate of 12byte per message with 140 messages per day daily capacity and that is very low [33].

Olamide Jogunola et al. present an adaptive routing algorithm technique based on distributed dual-gradient updates for facilitating communication among distributed energy resources units in smart grids. This algorithm can route messages across the network in fewer processing steps. Paper also showing potential cost-saving and reduced complexity which can save a significant amount of time for large system applications [34]. C.M. Colson and M.H. Nehrir present a method for facilitating the primary self-assembling and concerted departments within the microgrid agents. This paper clearly describes that decentralized multi-agent system methods are a viable alternative and deserve investigation, especially considering the daunting microgrid control problem [35].

Mohamed Salah Elbana et al. proposes the smart conversion of microgrid subsystems. The smart conversion includes a smart protection system, smart control algorithms with the ability to track fast system dynamics. The status analyzing approach is presented in accordance with the continuous synchronized measurements generated by μPMUs. To Identify, normal and fault cases, several algorithms are also presented [36].

Modern technology is works on Cyber-Physical System; this contains all users' data from the amount of electricity utilization to the location of the systems. Cybersecurity strategies are most essential for the implementation of the Modern grid [12]. Apparently, very few papers are published related to this topic [13].

The storage capacity available with Plug-in Electric Vehicle, which is in idle condition, that can be utilized for grid regulation activities. Also, in this paper issues and challenges due to the integration of heterogeneous entities like RES, Plug-in Electric Vehicle, in Micro Grid is discussed. Plug-in Electric Vehicle role and its usage as storage for grid ancillary support are discussed in detail. As a case study, the effective utilization of Plug-in Electric Vehicles to mitigate solar and load demand fluctuations is presented. Operational challenges pertaining to microgrid operations and smart home management are also presented [37].”

As per described in the passage, there has been many research is happening worldwide on the Smart grid technology; to analyze overall progress on these researches, some surveys and studies are carried out that addresses a wide range of issues as mentioned in context from the communication network to security issues [38-43]. However, the majority of the survey covers one or two aspects of the Smart grid and don’t cover the context of the Industries.

III. Future Direction

Smart Grid implementation in the Industrial zone is the biggest challenge due to the high-power network, costly production line, and sensible data. Even though it is necessary to test and investigate Smart Grid technology at the industrial sites for research and development. At the initial level, it is a proposal to establish a methodology for determining the optimal composition of a hybrid system for powering an industrial building.

Inputs:

1. Object load diagram with duration of 1 year (15 Min resolution).

2. Identification of types of renewable energy generation sources and batteries, most suitable for the building infrastructure.
3. Prices of hybrid system components (RESs: Euro/kW & Batteries: Euro/kWh).
4. Autonomous management system with mobile application.
5. Prices of energy from the utility (Euro/kWh)

Expected outputs:

1. Optimal performance of installed individual types of suitable RESs.
2. Optimal performance of installed energy storage device capacity
3. Participation of hybrid components to balancing the energy in a building.
4. The coverage of a given load course by the proposed hybrid system (Course of RESs production, Battery charging/Discharging, Course of energy consumption from Distribution Source)
5. Economic evolution of design (qualification of investment, operating cost, maintenance cost, comparison of saving by taking electricity without hybrid components and with hybrid components.)

The criteria are to identify minimum payback period time on investment (Maximum return with less investment).

A further possible step is to incorporate ESSs as common for 2-3 building infrastructure to optimize asset and financial performance, along with enhancing grid reliability and controllability.

Electric vehicles are also storage of electrical energy. Their charging point needs to set at the industrial site for G2V and V2G principal testing. The effect on frequency fluctuation needs to evaluate during the charging and discharging of EVs. Optimal performance condition could be identified after the practice.

Research on cyber security has to set on highest level for the protection of critical data. Block chain technology could be more effective in data transfer as HS codes get change with just small changes in the system. It would be faster to identify and secure technology than other most present technologies.

There are several technologies that must be carefully examined as an illustration of communication technology, automation technology and decentralized technology.

Government and Industries must come together for the deal with this expensive investment of the Smart grid. Policies and standard needs to be in favor of both utility companies and industrial consumers.

The final stage of testing is reliability and integration testing. The integration of most suitable devices according to availability in the present market, designing of a smart control panel with nanotechnology for saving space, that also inhomogeneous method with real-time data transferring process to identify most promising devices and fault location tracking. Smart Grid diagram of industrial sector is shown in Fig.1.

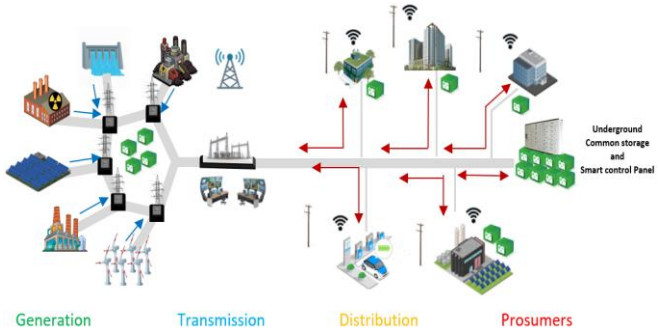


Fig. 1. Smart Grid Diagram of Industrial Zone

REFERENCES

[1] Choobineh, M., Mohagheghi, S., 2016. A multi-objective optimization framework for energy and asset management in an industrial Microgrid. J. Clean. Prod. 139, 1326e1338. https://doi.org/10.1016/j.jclepro.2016.08.138
[2] http://www.eia.doe.gov/emeu/mecs/mecs2006/pdf/Table11_1.pdf
[3] Tariq Samad, Sila Kiliccote, Smart grid technologies and applications for industrial sector, Honeywell automation and control solutions, Lawrence Berkeley National Laboratory, Savannah, GA, 10 Jan. 2012
[4] Samad, T.; Kiliccote, S. Smart grid technologies and applications for the industrial sector. Comput. Chem. Eng. 2012, 47, 76–84.
[5] European Environment Agency (EEA). Achieving Energy Efficiency through Behaviour Change: What Does it Take? European Union: Brussels, Belgium, 2013.
[6] Gangale, F.; Mengolini, A.; Onyeji, I. Consumer engagement: An insight from smart grid projects in Europe. Energy Policy 2013, 60, 621–628.
[7] ABB| Industrial Smart Grid, 3AFA 701889 / 10.2.2011
[8] Xenos, D.P., Mohd Noor, I., Matloubi, M., Cicciotti, M., Haugen, T., Thornhill, N.F., 2016. Demand-side management and optimal operation of industrial electricity consumers: an example of an energy-intensive chemical plant. Appl. Energy 182, 418e433. https://doi.org/10.1016/j.apenergy.2016.08.084.
[9] Schaab, D.A., Weckmann, S., Kuhlmann, T., Sauer, A., 2018. Simulative analysis of a flexible, robust and sustainable energy supply through industrial smart-DC-grid with distributed grid management. Procedia CIRP 69, 366e370. https://doi.org/ 10.1016/j.procir.2017.11.037.
[10] Schulze, M., Nehler, H., Ottosson, M., Thollander, P., 2016. Energy management in industry - a systematic review of previous findings and an integrative conceptual framework. J. Clean. Prod. 112, 3692e3708. https://doi.org/10.1016/j.jclepro.2015.06.060.
[11] Ma, Z., Jorgensen, B.N., Asmussen, A., 2015. Industrial consumers' acceptance to the smart grid solutions: case studies from Denmark. In: Proc. 2015 IEEE Innov. Smart Grid Technol. - Asia, ISGT ASIA 2015.
[12] Oyetoyan, T.D., Conradi, R., Sand, K., 2012. Initial survey of smart grid activities in the Norwegian energy sector - use cases, industrial challenges and implications for research. In: 2012 1st Int. Work. Softw. Eng. Challenges Smart Grid, SESmartGrids 2012 - Proc, vol. 2, pp. 34e37.
[13] Alexander Brem, Mumtahina Mahajabin Adrita, Dominic T.J. O'Sullivan, Ken Bruton, Industrial smart and micro grid systems e A systematic mapping study, Department of Mechanical, Biomedical and Manufacturing Engineering, Cork Institute of Technology, Bishopstown, Cork, Ireland, Intelligent Efficiency Research Group (IERG), Department of Civil and Environmental Engineering, University College Cork, College Road, Cork, Ireland, https://doi.org/10.1016/j.jclepro.2019.118828
[14] Ancillotti, E., Bruno, R., Conti, M., 2013. The role of communication systems in smart grids: architectures, technical solutions and research challenges. Comput. Commun. 36, 1665e1697. https://doi.org/10.1016/j.comcom.2013.09.004.
[15] Teive, R.C.G., Andrade, F.F., Aranha Neto, E.A.C., Ros_ario, L.M., De Bettio, J.A., 2015. Novel method for typical load curves characterization of industrial consumers towards the smart grids. In: 2015 18th Int. Conf. Intell. Syst. Appl. To Power Syst. ISAP 2015, pp. 1e7.
[16] Noor Ahmed Qarabsh, Sana Sabah Sabry, Haneen Ahmed Qarabash , Smart grid in the context of industry 4.0: an overview of communications technologies and challenges, Indonesian Journal of

Electrical Engineering and Computer Science, College of Engineering, University Of Information Technology and Communications, College of science, University of Baghdad, Vol. 18, No. 2, May 2020, pp. 656~665

[17] Gamarra, C., Guerrero, J.M., Montero, E., 2016. A knowledge discovery in databases approach for industrial microgrid planning. Renew. Sustain. Energy Rev. 60, 615e630. https://doi.org/10.1016/j.rser.2016.01.091.

[18] Most efficient solar panels: solar panel cell efficiency explained, Energysage, https://news.energysage.com/what-are-the-most-efficient-solar-panels-on-the market/#:~:text=The%20most%20efficient%20solar%20panels%20on%20the%20market%20today%20have,brand%20available%20on%20the%20market.

[19] Flow battery. Retrieved from Wiki pedia: https://en.wikipedia.org/wiki/Flow_battery

[20] Storage Solution. Retrieved from Energy Vault: https://energyvault.com/#problem-solution

[21] Project Profile: High-Efficiency Thermal Storage System for Solar Plants. Retrieved from www.energy.gov: https://www.energy.gov/eere/solar/project-profile-high-efficiency-thermal-storage-system-solar-plants

[22] Hydrostor. Retrieved from Hydrostor: https://www.hydrostor.ca/wp-content/uploads/2020/01/Hydrostor_Brochure_2020.pdf

[23] Cryogenic energy storage. Retrieved from highviewpower: https://www.highviewpower.com/technology/

[24] The Ambri technology. Retrieved from Ambri: https://ambri.com/technology/

[25] Solid-state battery. (2019, July). Retrieved from Wiki Pedia: https://en.wikipedia.org/wiki/Solid-state_battery#Advantages

[26] William Sanchez-Huertas1, Víctor Gómez1, Cesar Hernández, Optimization Algorithms for Solving Microgrid and Smart Grid Integration Problems, Universidad Distrital Francisco José de Caldas, Technological Faculty, Bogotá, Colombia, Corresponding author, 1*Orcid: 0000-0001-9409-8341

[27] H. Yoo and T. Shon, "Challenges and research directions for heterogeneous cyber–physical system based on IEC 61850: Vulnerabilities, security requirements, and security architecture," Future generation computer systems, vol. 61, pp. 128-136, 2016.

[28] N. Andreadou, M. Guardiola, and G. Fulli, "Telecommunication technologies for smart grid projects with focus on smart metering applications," Energies, vol. 9, p. 375, 2016.

[29] D. Baimel, S. Tapuchi, and N. Baimel, "Smart grid communication technologies-overview, research challenges and opportunities," in Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2016 International Symposium on, 2016, pp. 116-120.

[30] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "A survey of 5G technologies: regulatory, standardization and industrial perspectives," Digital Communications and Networks, vol. 4, pp. 87-97, 2018.

[31] G. Shi and K. Li, "Fundamentals of ZigBee and WIFI," in Signal Interference in WiFi and ZigBee Networks, ed: Springer, 2017, pp. 9-27.

[32] D. G. Kogias, E. T. Michailidis, G. Tuna, and V. C. Gungor, "Realizing the Wireless Technology in Internet of Things (IoT)," in Emerging Wireless Communication and Network Technologies, ed: Springer, 2018, pp. 173-192.

[33] P. Sethi and S. R. Sarangi, "Internet of things: architectures, protocols, and applications," Journal of Electrical and Computer Engineering, vol. 2017, 2017.

[34] Jogunola _ Bamidele Adebisi _ Kelvin Anoh _ Augustine Ikpehai __ Mohammad Hammoudeh, Adaptive Routing Algorithm for Information Management in Distributed Microgrids in Smart Grid Olamide, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M1 5GD, UK, Department of Engineering and Mathematics, Sheffield Hallam University, Sheffield. S1 1WB, UK

[35] C.M. Colson, Student Member, IEEE, M.H. Nehrir, Fellow, IEEE, Algorithms for Distributed Decision-Making for Multi-agent Microgrid Power Management.

[36] Mohamed salah elbana, nabil h. Abassy, ashraf meghad, Smart control algorithms for microgrid status Detection, optimization and protection, MSc student, Ph.D, Faculty of Engineering, Alexandria University, Egypt

[37] K. Ramakrishna Reddy and S. Meikandasivam, Smart Distribution Network with Integration of Stochastic Renewable Energy Sources and Plug-in Electric Vehicles: Challenges and Issues, Vellore Institute of Technology, Vellore, India Received 05 May 2018; Accepted 17 September 2018, Publication 10 December 2018

[38] D. Zhang, C. C. Chan, and G. Y. Zhou, "Enabling Industrial Internet of Things (IIoT) towards an emerging smart energy system," Global Energy Interconnection, vol. 1, pp. 39-47, 2018.

[39] M. Moretti, S. N. Djomo, H. Azadi, K. May, K. De Vos, S. Van Passel, et al., "A systematic review of environmental and economic impacts of smart grids," Renewable and Sustainable Energy Reviews, vol. 68, pp. 888-898, 2017.

[40] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, et al., "A survey on smart grid potential applications and communication requirements," IEEE Transactions on industrial informatics, vol. 9, pp. 28-42, 2012.

[41] S. S. Reka and T. Dragicevic, "Future effectual role of energy delivery: A comprehensive review of Internet of Things and smart grid," Renewable and Sustainable Energy Reviews, vol. 91, pp. 90-108, 2018.

[42] N. S. Nafi, K. Ahmed, M. A. Gregory, and M. Datta, "A survey of smart grid architectures, applications, benefits and standardization," Journal of Network and Computer Applications, vol. 76, pp. 23-36, 2016.

[43] M. Ourahou, W. Ayrir, B. E. Hassouni, and A. Haddi, "Review on smart grid control and reliability in presence of renewable energies: Challenges and prospects," Mathematics and Computers in Simulation, 2018.

Lung and Pleura Segmentation in Lung Ultrasound with Deep Learning

[1]*Miroslav Jaščur (4^{th} year),*
Supervisor: [2] *Marek BUNDZEL*

[1,2]Department of Cybernetics and Artificial Intelligence

[1]miroslav.jascur@tuke.sk, [2]marek.bundzel@tuke.sk

***Abstract*—Particular post thoracic surgery complications are standardly monitored using methods that employ ionizing radiation. Following a clinical trial of novel lung ultrasound (LUS) examination procedure that can replace X-raying, a need to automatize the diagnosis procedure has arisen. Deep learning was used as a powerful tool for LUS analysis. We present a novel deep learning method - Motion Mode (M-Mode) classification - to detect lung sliding motion absence in LUS. One of the modules in M-Mode classification is semantic segmentation module. In this paper, we present our semantic segmentation results. We use convolutional neural network, U-Net, to segment lung and pleura tissue in LUS. Our segmentation accuracy was 98.63% and 92.92%, for the pleura and for the lung, respectively. While the pleura segmentation U-Net achieves higher accuracy, this performance difference does not translate into the metric intersection over union (IoU), where the lung segmentation U-Net outperforms the pleura segmentation U-Net by 17.38% and achieves IoU of 75.18%.**

***Keywords*—Deep Learning, Lung Ultrasound, Semantic Segmentation, U-Net.**

I. Introduction

Feature extraction from ultrasound (US) heavily relies on the radiologist's expertise, and this may be difficult to share, especially if a novel diagnostic protocol is being implemented, such as in our case. Automated computer vision methods and machine learning algorithms that can distill information from US images or videos have been introduced over the past decades. DL methods can transform raw image data to a high level of abstraction and automate feature extraction from the US. They can be applied to image analysis and aid in decision support systems. To our best knowledge, DL has not been applied to analyze LUS in the context of the detection of post thoracic surgery complications. One of the crucial preprocessing steps in the lung ultrasound analysis is semantic segmentation.

Semantic image segmentation is the task of assigning a class label to every pixel in an image. To perform quantitative analysis of pathological and non-pathological tissues, we often need to locate them in a US image. Fully convolutional neural networks such as U-Nets [1] are state-of-the-art in biomedical image processing. For the segmentation of rigid structures, fine-tuned CNNs paired with an effective post-processing method are currently the best-performing approaches [2], [3]. For the segmentation of non-rigid anatomical structures, prior information is often introduced to the process. This information can be provided through active contours, deformable templates or image registration.

A. Segmentation Labels

A highly trained expert must label LUS video files, but the procedure is not time consuming. On the other hand, labeling a 2D dataset for basic semantic segmentation in LUS is more time expensive but requires only low expertise. We have educated a group of 80 volunteers, who have drawn rectangular bounding boxes around the objects of interest using labelImg tool [4]. Labeling was the first step in the production of semantic segmentation masks.

We randomly selected 6400 frames from 182 videos, and each volunteer labeled 80 frames. The objects of interest were: *pleura*, *lung*, *ribs* and *tissue*. Class *pleura* represents both the parietal and visceral pleura. In the LUS image, it is a shimmering white line connecting the lung and the chest wall. Class *lung* is the area under the pleura between the anechoic shadows of the ribs. Class *ribs* represents the area of anechoic shadow. This artifact appears because the sound waves are well reflected by the hard material of the ribs. Finally, the class *tissue* represents the chest wall comprised of the skin, fat, and muscle tissue.

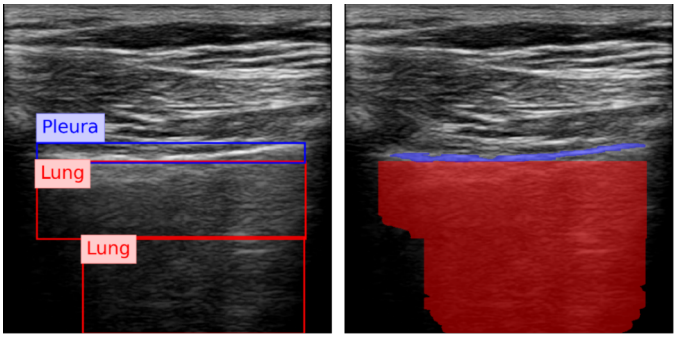


Fig. 1. Creation of semantic masks for pleura and lung from labeled data. On the left: bounding boxes labeled by volunteers. On the right: semantic masks generated from the bounding boxes.

We produced binary semantic segmentation masks for the objects of interest by applying several computer vision algorithms to the areas delimited by the bounding boxes. We assumed that the objects of interest differ from other objects in the bounding box by brightness. To create the segmentation mask for the *lung* class, see Fig. 1, we applied intensity thresholding in the bounding box with a threshold value determined by Otsu's method. The initial result was refined by morphological dilation, and the holes in the mask were filled. Finally, the mask was returned to the coordinate system

of the image. We repeated this process for each bounding box in an image. In the end, we kept only the largest contiguous region because some of the images contained multiple separate regions of the lung. We generated *pleura* masks using the same process, see Fig. 1.

II. Lung and pleura segmentation

For the segmentation of the lung and pleura we used fully convolutional architecture - U-Net [1]. Our model contains three encoder and three decoder blocks. Each block uses a 2D convolutional layer, 2D batch normalization, and a rectified linear unit layer in this order, twice, followed by either 2D max-pooling in the case of the encoder block or 2D transposed convolution in the case of the decoder block. The first two encoder blocks connect to a deeper layer and skip-connect to the corresponding decoder layer with the same resolution. The third and deepest encoder block connects directly to the deepest decoder block. Decoder blocks take the skip connected output from the encoder block and the output from the decoder's deeper layer and produce the input to the higher resolution decoder block. Finally, the prediction is made using a softmax layer. Loss is calculated as weighted pixelwise cross-entropy. The weight is determined as a class distribution from a random subsample of the dataset. In the pleura segmentation, decreasing the weight of pleura segments by 80% improved the segmentation results.

The input image resolution is 480×480 pixels. We used four types of data augmentation: vertical and horizontal flipping and random rotation, translation, and scaling. We trained the neural networks for 30 epochs with early stopping that occurred when after ten epochs, the validation loss did not drop below the lowest previously recorded value. We randomly divided the dataset into three parts: 50% for training and 25% each, for validation and testing.

To improve the resulting segmentation, we apply post-processing both for the pleura and for the lung:

- The connected regions in the prediction mask are detected.
- The largest connected region is kept, the remaining are discarded.
- The holes in the largest connected region are filled.

III. Evaluation

We use accuracy, precision, recall, and F1-score to evaluate both semantic segmentation and classification CNNs. We add Intersection over Union (IoU) metric to evaluate semantic segmentation results and use the term Dice coefficient instead of F1-score.

In binary segmentation, the tissue (lung/pleura) is our target class. Therefore, accuracy represents the ratio of the correctly classified pixels to all pixels in the mask. Precision measures the proportion of the correctly classified tissue pixels to all pixels classified as tissue. Recall measures the proportion of correctly classified background pixels to all pixels classified as background. The dice coefficient is a harmonic mean of precision and recall. IoU measures the overlap between the correctly identified tissue pixels and the area of the union of all pixels identified as tissue pixels and misclassified background pixels.

IV. Segmentation results

Tab. I shows the performance metrics of U-Nets' binary segmentation of the pleura and the lung on the test dataset. We use accuracy, recall, precision, IoU, and Dice coefficient. In segmentation, improving IoU metrics is the criterion.

TABLE I
U-Net performance characteristics for the segmentation of pleura and lung (Post.- indicates if post-processing according to Sec. II was applied. Acc. - accuracy, IoU - Intersection over Union, Dice - Dice Coefficient.)

Tissue	Post.	Acc.	Precision	Recall	IoU	Dice
Pleura	No	0.9816	0.7580	0.5985	0.5025	0.6689
Pleura	Yes	0.9863	0.7497	0.7093	**0.5735**	0.7289
Lung	No	0.9198	0.9390	0.7644	0.7283	0.8428
Lung	Yes	0.9292	0.9369	0.7919	**0.7518**	0.8584

Without post-processing, the pleura segmentation was 96.81% accurate, but the IoU score was only 50.25%. By applying post-processing described in Section II, the resulting accuracy improved by 0.47% and IoU by 7.1%. The post-processing improvement was at the cost of decreased precision caused by the removal of true positive pixels.

Lung segmentation with U-Net was more successful than pleura segmentation with an IoU higher by 32% without post-processing. The overall accuracy is lower with or without post-processing because the lung samples are pixel-wise more balanced than the pleura samples. With post-processing, every metric improved except for precision that decreased by 0.21%. The best-performing classifier achieved an accuracy of 92.92% and IoU 75.18%.

V. Conclusion

The U-Net for lung segmentation outperforms the U-Net for pleura segmentation in almost every measure, see Tab. I. These performance differences result from different tissue to background ratios and the possible mislabeling during the process of semi-automated production of semantic segmentation masks. Furthermore, the lung tissue is constrained by the pleura and the anechoic shadows of the ribs, but the contours of the pleura tissue are more fuzzy.

The full version of the article and results of current research are under review. Other publications of the author include [5], [6].

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[2] Y. Zhang, M. T. Ying, L. Yang, A. T. Ahuja, and D. Z. Chen, "Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2016, pp. 443–448.

[3] Y. Li, R. Xu, J. Ohya, and H. Iwata, "Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 1485–1488.

[4] D. Tzutalin, "Labelimg," https://github.com/tzutalin/labelImg, 2020.

[5] M. Bundzel, M. Jaščur, M. Kováč, T. Lieskovskỳ, P. Sinčák, and T. Tkáčik, "Semantic segmentation of airborne lidar data in maya archaeology," *Remote Sensing*, vol. 12, no. 22, p. 3685, 2020.

[6] N. Ferenčík, M. Jaščur, M. Bundzel, and F. Cavallo, "The rehapiano—detecting, measuring, and analyzing action tremor using strain gauges," *Sensors*, vol. 20, no. 3, p. 663, 2020.

Author's Index

A
Adjailia Fouzia 42
Alexovič Stanislav 75

B
Bačkai Július 87
Baumgartner Maroš 133
Biľanský Juraj 83
Buday Anton 143
Bugata Peter 61

Č
Čík Ivan 157

D
Durkáč Martin 186
Dzivý Daniel 121

F
Fecko Branislav 63
Fedor Marek 35

G
Gazda Matej 93
Gecášek Daniel 91
Gereg Slavomír 119
Gurbáľ Filip 139

H
Haluška Renát 192
Hasin Martin 115
Havran Peter 16
Havrilla Martin 111
Hireš Máté 59
Horniaková Jana 167
Hrabovský Peter 171
Hruška Lukáš 153
Hudák Marián 67
Hulič Michal 145
Humeník Jozef 95

I
Ivan Jozef 85

J
Jaščur Miroslav 199

K
Karpets Maksym 125
Koctúrová Marianna 151
Kohan Vladimír 30
Kolárik Michal 97
Koska Lukáš 181
Kuzmiak Marek 39

L
Lapčák Maroš 44
Lenger Tomáš 169

M
Madeja Matej 10
Magyar Ján 12
Marcinek Adrián 160
Margitová Anastázia 27
Martinko Dávid 65
Maslej Krešňáková Viera 137

O
Olexa Richard 183
Oliinyk Maksym 123

P
Palša Jakub 14
Pavlišková Viera 105
Pál Daniel 21
Pella Zuzana 102
Petija Rastislav 18
Petro Viktor 147
Pomšár Ladislav 155
Pusztová Ľudmila 109

R
Rasamoelina A. David 173
Ružička Marek 32

S
Samuhel Simeon 79
Sivý Martin 131
Sokol Miroslav 178
Solanik Michal 127

Š
Šárpataky Ľuboš 54
Šárpataky Miloš 48
Štefko Róbert 23
Šuľaj Peter 190

T
Tailor Rikin 194
Tkáčik Milan 176

U
Urbanský Jakub 69

V
Varga Michal 165

Z
Zdravecký Norbert 71

SCYR 2021: 21st Scientific Conference of Young Researchers

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
Edition I, 202 pages
Number of CD Proceedings: 50 pieces

Editors: prof. Ing. Alena Pietriková, CSc.
Ing. Emília Pietriková, PhD.

ISBN 978-80-553-3904-7